Course Notes on Computational Finance

Rüdiger Seydel

Universität zu Köln
Mathematisches Institut

2015

Version 2015

These course notes can be downloaded from
http://www.mi.uni-koeln.de/∼seydel/skript.html

seydel@mi.uni-koeln.de
www.compfin.de

CCFinance
Cologne Computational Finance Skriptum

Preface

These are the course notes of my introductory lectures on Computational Finance. This course is a one-semester course, 4 hours per week, with exercises as homework. The level is tailored to undergraduate students.

The choice of subjects provides an elementary introduction into the field. Basic knowledge on financial markets is assumed, but otherwise no specific prerequisites are needed, except for an education in calculus and basic numerical analysis. The style is expository. Ideas are put forward, and the methods are summarized as algorithms. Numerical experiments and work on the exercises are essential for understanding.

The author has developed this course over the past 15 years, experimenting with different kind of explanations and expositions. Rich experience in teaching the material has flown into my book *Tools for Computational Finance*; in some way these course notes may be seen as a condensed version of essential parts.

Köln, July 2014 Rüdiger Seydel

Contents

0. (Prologue) Financial Derivatives

A. Overview

Assets include

equities, stocks

bonds

commodities

spot market: delivery on the same date

derivatives markets:

contract today $(t = 0)$

settlement at a future date (i.e. for $t = T > 0$, T maturity, in years)

Notation: S_t (or $S(t)$) is the market price of the asset at time t.

Examples of risks

1.) A stock holder wishes to sell 1000 shares in 6 months. Today's spot price per share is 50\$. *Risk*: *Crash*! Aim: in 6 months these shares should be worth at least $50,000$\$ $(T = 0.5)$.

2.) A farmer expects in three months (in July) a crop of 800 tons of corn. *Risks*: crop failure, or price decline. Aim: sell the crop today, and deliver end of July $(T = 0.25)$. Question: which price is adequate?

3.) Buy today a shipment of oil (in Dollar) supposed to arrive in two weeks in Rotterdam, where it is supposed to be sold (in Euro). *Risks*: delayed arrival, changes in price and exchange rate.

4.) A company raises a loan with variable interest rate $r_v(t)$. *Risk*: increasing rate.

and so on.

Problem in each case: The crucial future value S_T is unknown today $(t = 0)$. Known is S_0 and the "current" interest rate, and possibly further information from the past $(t \leq 0)$.

But:
The contract is settled today and fixes a price F at maturity T.

Question:
What is the fair price that enables an agreement of both parties (seller and buyer)? This main question goes along with other questions, for example, how will "the interest" rate develop?

Most of those risks can be hedged with financial **derivatives**, which are instruments derived from underlying assets. Three kind of derivatives are

futures/forwards
options
swaps

Our main concern: What is a reasonable price of such derivatives?

"Financial Engineering"

For the pricing of futures/forwards typically simple computational tool suffice. But options require sophisticated numerical algorithms.

This course: essentially numerical methods for options. Require knowledge on

foundations of stochastics,
differential equations, including partial differential equations (PDEs)
numerical analysis

Contents

B. Forwards and Arbitrage

Definition: A **forward** is a contract between two parties A and B to buy or sell an asset at a future time t such that A delivers the asset (spot price today S_0), and B pays the price F.

Proposition:

When for $0 \le t \le T$ the asset does not produce income nor costs, then

$$F = S_0 \mathrm{e}^{rT}$$

Here r is the interest rate of a riskless bond with time to maturity T, and two technical assumptions hold (see below). The proof follows from the no-arbitrage principle.

Definition: Arbitrage = risk-free profit

or more precise:

> arbitrage means the existence of a portfolio, which requires no investment initially, and which makes no loss but very likely a gain at maturity.

In a (fictitious) idealized market, information spreads rapidly and arbitrage opportunities become apparent and do not last long.

General assumption: There is no arbitrage! (*no-arbitrage principle*)

further assumptions:

1.) continuously compounded interest (factor e^{rT})

2.) lending rate and borrowing rate are equal.

Proof of $\mathbf{F = S_0 e^{rT}}$:

1. Assume $F > S_0\mathrm{e}^{rT}$.
   Then there is an arbitrage-strategy as follows:
   $t = 0$: borrow S_0 at the interest rate r,
   buy the asset,
   enter a forward contract, to sell the asset at $t = T$ for the price F.
   $t = T$: sell the asset: $+F$
   repay the loan $-S_0\mathrm{e}^{rT}$
   result: $F - S_0\mathrm{e}^{rT} > 0$,
   which is a riskless profit, and hence arbitrage!
   This contradicts the no-arbitrage principle $\Rightarrow\ F \le S_0\mathrm{e}^{rT}$
2. Assume $F < S_0\mathrm{e}^{rT}$.
   $t = 0$: Investors who own the asset sell it: $+S_0$
   invest S_0 at interest rate r,
   enter a forward-contract to buy the asset at price F
   $t = T$: from the contract: $-F$
   from the investment: $+ S_0\mathrm{e}^{rT}$
   result $S_0\mathrm{e}^{rT} - F > 0$ hence arbitrage
   contradiction! $\Rightarrow$ $F \ge S_0\mathrm{e}^{rT}$
3. together: $F = S_0\mathrm{e}^{rT}$

Remark: For two interest rates r_1, r_2 obtain $F \le S_0\mathrm{e}^{r_1 T}$; $F \ge S_0\mathrm{e}^{r_2 T}$;
hence: $S_0\mathrm{e}^{r_1 T} \le F \le S_0\mathrm{e}^{r_2 T}$

C. Interest Rate

At time t, lend money $N(t)$ for the short period Δt. Let $r(t, \Delta t)$ be the corresponding interest rate,

$$N(t + \Delta t) - N(t) = r(t, \Delta t)\, N(t)\, \Delta t\,.$$

The discount factor for the interest period from t to T is

$$Z_r(t, T) := \frac{N(t)}{N(T)}$$

For time-continuous $t \in \mathbb{R}$ and integrable r follows:

$$N(T) = N(t) \, \exp \int_t^T r(x) \, \mathrm{d}x \, .$$

Specifically for constant r we have

$$Z(t,T) = \mathrm{e}^{-r(T-t)} \, .$$

risk-free interest rate: The rate earned on a riskless asset is the *risk-free* interest rate. Take zero-bonds for its derivation out of market data.

For the following, r always means the risk-free interest rate that matches the time horizon $T - t$.

D. Literature

The standard literature for this course is:

R.U. Seydel: Tools for Computational Finance. 6th Ed., Springer, London 2017

For financial background:

J.C. Hull: Options, Futures, and Other Derivates. Prentice Hall.

For colored illustrations and further explanation we recommend to consult

Topics in Computational Finance

This collection is part of the homepage *www.compfin.de*

1. Modeling of Financial Options

1.1 Options

Definition (Option)

An *option* is the right (but not the obligation) to buy or sell a risky asset at a prespecified fixed "strike" price K until a maturity time T.

The terms of the option contract are fixed by the *writer*. The *holder* of the option pays a premium V for its purchase.

Exercising the option means to buy or sell the *underlying* asset for the price K according to the option's contract. An option with the right to buy the underlying is called *call*, and the option to sell is called *put*.

Question: What is the fair premium V?

This depends on the price K, on the price S_0, on T, and on market data such as the rate r or the volatility σ.

The *volatility* σ measures the size of fluctuations of the asset price S_t, and hence indicates the risk.

A *European* option can only be exercised at maturity ($t = T$); an American option can be exercised anytime during the life time $0 \le t \le T$.

The value of the premium V at maturity is easy to assess: it is the *payoff*.

1. Call in $t = T$

The holder of the option has two alternatives to acquire the asset:

(a) She buys it on the spot market and pays S_T, or

(b) exercises the call option and pays the strike price K.

The rational holder optimizes her position.

1st case: $S_T \le K \Rightarrow$ The holder pays S_T on the spot market, and lets the option expire. Then the option is worthless, $V = 0$.

2nd case: $S_T > K \Rightarrow$ The holder exercises the call and pays K. And immediately she sells the asset for the spot price S_T. The profit is $S_T - K$, hence $V = S_T - K$.

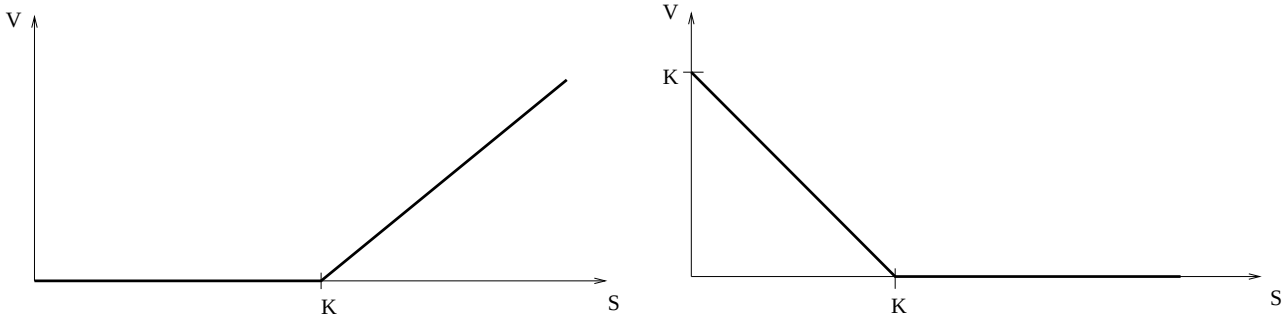
In summary, the payoff of a call is

$$V(S_T, T) = \begin{cases} 0 & \text{in case } S_T \le K \\ S_T - K & \text{in case } S_T > K \end{cases}$$
$$= \max\{S_T - K, 0\} \quad =: (S_T - K)^+$$

2. Put in $t = T$

Analogous reasoning leads to the payoff of a put:

$$V(S_T, T) = \begin{cases} K - S_T & \text{in case } S_T \le K \\ 0 & \text{in case } S_T > K \end{cases}$$
$$= \max\{K - S_T, 0\} \quad =: (K - S_T)^+$$



Payoff of a call (left) and of a put (right), in $t = T$.

The same arguing is valid for American-style options for any $t \le T$: the payoffs are

$$\begin{aligned} \text{put:} \quad & (K - S_t)^+ \\ \text{call:} \quad & (S_t - K)^+ \end{aligned}$$

The value V for $t < T$, in particular for $t = 0$, is more difficult to determine. The *no-arbitrage-principle* plays a central role. This mere principle leads to **bounds** for V. We give some examples.

The value $V(S, t)$ of an American option can not be smaller than the payoff, because (proof for a put; call is analogous):

> Obviously $V \ge 0$ for all S. *Assume*: $S < K$ and $0 \le V < K - S$. Establish arbitrage as follows: Buy the asset $(-S)$ and the put $(-V)$, and exercise immediately: $(+K)$. By $K > S + V$ this is a risk-free profit $K - S - V > 0$, which contradicts the no-arbitrage-principle.

Hence

$$V^{\text{Am}}_{\text{Put}}(S, t) \ge (K - S)^+ \quad \forall S, t\,.$$

Analogously:

$$V^{\text{Am}}_{\text{Call}}(S, t) \ge (S - K)^+ \quad \forall S, t\,.$$

Also the inequality

$$V^{\text{Am}} \ge V^{\text{Eu}}$$

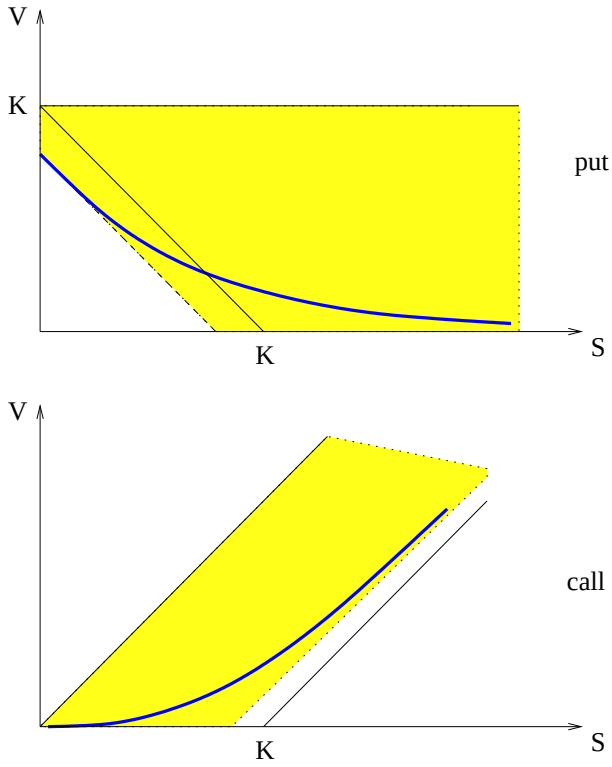
holds since an American option embraces the European option. When no dividend is paid, the put-call parity

$$S + V_{\text{Put}} = V_{\text{Call}} + K\mathrm{e}^{-r(T-t)}$$

holds for European-style options. This leads to further bounds, for example, to

$$V^{\text{Eu}}_{\text{Put}} \ge K\mathrm{e}^{-r(T-t)} - S.$$

The figure illustrates the a-priori bounds for European options on assets that pay no dividends for $0 \le t \le T$ (for $r > 0$).



Definition (historic volatility)

The historic volatility σ is the standard deviation of S_t. It is scaled by $\frac{1}{\sqrt{\Delta t}}$ since the data are returns sampled at Δt. In reality, σ is not constant, but the classic Black–Scholes-model takes it as constant. The empirical determination of market parameters (such as σ) is an ambitious task (*calibration*).

Notice that each option involves three prices, namely, the price S_t of the underlying asset, the strike price K and the premium V of the option.

Definition

Options with the above payoffs $\Psi(S) := (K - S)^+$ or $\Psi(S) := (S - K)^+$ on a single asset are called *standard options*, or *vanilla options*. There are many other kinds of options with other features. These other types of options are called *exotic*.

Examples of exotic options

Basket: The underlying is a basket of several assets, e.g., $\sum_{i=1}^{m} w_i S_i(t)$, where S_i is the market price of the ith asset, $m > 1$.

Options with *other payoffs*, such as the *binary* put with

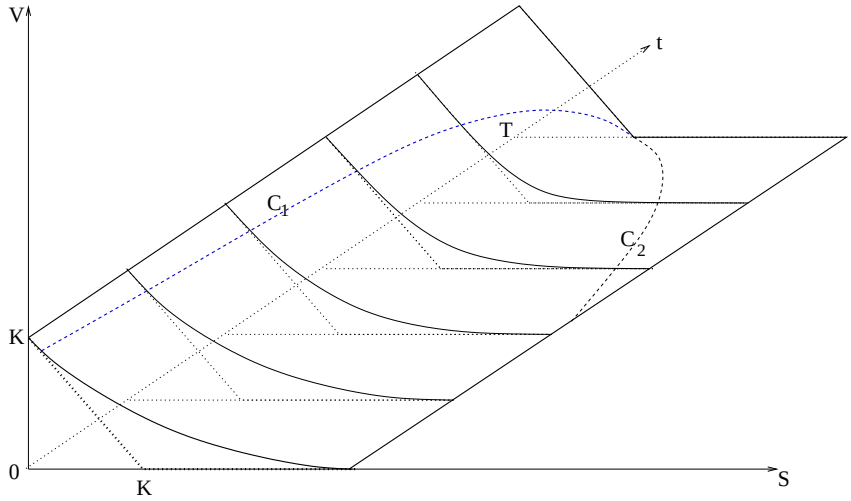
$$\text{payoff} = \begin{cases} 0 & \text{in case } S_t > K \\ 1 & \text{in case } S_t \le K. \end{cases}$$

Path dependence: For instance, the payoff $(\frac{1}{T}\int_0^T S(t)\,\mathrm{d}t - K)^+$ involves the average value, which depends on the path of $S(t)$ (*average price call*).

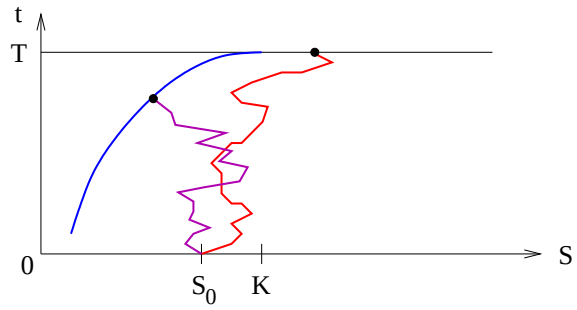
Barrier: For instance, an option ceases to exist when S_t reaches a prespecified barrier B.

On the Geometry of options

The values $V(S,t)$ obey the bounds sketched above, see the illustration of an American put. $V(S,t)$ can be interpreted as surface over the half strip $0 \le t \le T$, $S > 0$. This $V(S,t)$ is called *value function*. At the *early-exercise* curve, the surface merges in the plane defined by the payoff.



Importance: When the market price S_t reaches this curve C_1, immediate exercise is optimal: invest K for the interest rate r. The situation is sketched in an (S,t)-plane for an American put that pays no dividend.



For American call options *with* dividend payment the situation is analogous. The geometry at early-exercise curves will be discussed in Chapter 4. The curve must be calculated numerically.

1.2 Mathematical Model

A. Black–Scholes Market

Here we discuss mathematical models of how paths S_t may behave. We list some assumptions, which essentially go back to Black, Scholes and Merton (1973, Nobel-Prize 1997). These classic assumptions lead to a partial differential equation (PDE), the famous Black–

Scholes equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\,\sigma^2\,S^2\,\frac{\partial^2 V}{\partial S^2} + r\,S\,\frac{\partial V}{\partial S} - r\,V = 0$$

This equation is a symbol representing the classic theory. Each solution $V(S,t)$ of a European standard option must solve this PDE, satisfying for $t = T$ the terminal condition $V(S,T) = \Psi(S)$ where Ψ denotes the payoff.

Assumptions of the Model

1. There is no arbitrage.
2. The market is frictionless. That is, there are no transaction costs, and rates for lending and borrowing money are equal. All variables are perfectly divisible ($\in \mathbb{R}$). And individual trading does affect the market price.
3. The price S_t follows a geometric Brownian motion (explained later).
4. Technical assumptions:

   r and σ are constant for $0 \le t \le T$. No dividends are paid in $0 \le t \le T$.

Provided these assumptions hold (some can be weakened), **the value function of a European standard option solves the Black–Scholes equation**. Hence a possible approach to price a European option is to solve the Black–Scholes equation. There is an analytical solution; this is given at the end of this chapter (with δ a continuous dividend rate).

The above model of a finance market is the classic approach; there are other market models.

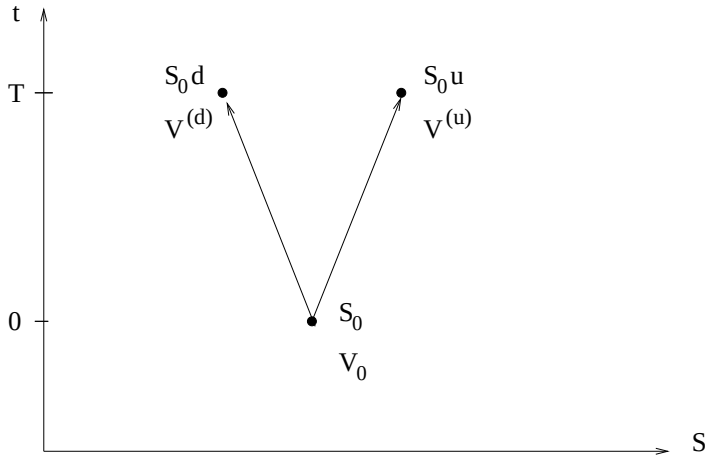
The model with its geometric Brownian motion (Section 1.5) is a continuous-time model, $t \in \mathbb{R}$. There are also discrete-time models, which consider only discrete time instances. Other market models do not use the geometric Brownian motion. Such models are mainly working with jump processes.

Numerical Tasks:

- Computation of $V(S,t)$, in particular for $t = 0$, with early-exercise curve for American options,
- Computation of sensitivities ("Greeks"), such as $\frac{\partial V(S,0)}{\partial S}$,
- Calibration, which means to estimate parameters that match empirical data.

B. Risk-Neutral Probabilities (One-Period Model)

Assumptions: $0 < d < u$, and the situation of the figure below. There are only two time instances: $0, T$, and two possible future asset prices $S_0 d$, $S_0 u$. V_0 denotes the (unknown) value of the option "today" for $t = 0$, and S_0 is the current value of the asset.



Consider a portfolio with two positions:

1. Δ shares of the asset
2. a short position of one option written on this asset

With Π_t denoting the wealth function, the value Π_0 of the portfolio at the time 0 is

$$\Pi_0 = S_0\,\Delta - V_0\,.$$

The number Δ is to be determined. At time T the value of the underlying is "up" or "down" and the portfolio is

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

$V^{(u)}$ and $V^{(d)}$ are fixed by the payoff. Choose Δ such that the portfolio becomes riskless at time T. That is, the value of the portfolio should be the same, no matter whether the market price goes "up" or "down",

$$\Pi^{(u)} = \Pi^{(d)} =: \Pi_T.$$

Consequently,

$$S_0\Delta(u-d) = V^{(u)} - V^{(d)}$$
$$\text{or} \quad \Delta = \frac{V^{(u)} - V^{(d)}}{S_0 u - S_0 d}.$$

With this special value of Δ the portfolio is riskless. Invoking the no-arbitrage principle, we conclude: Any other risk-free investment must have the same value, because otherwise arbitrageurs would make a riskless profit by exchanging the investments. Hence: $\Pi_T = \Pi_0\ \mathrm{e}^{rT}$

An elementary calculation shows

$$V_0 = \mathrm{e}^{-rT}\left(V^{(u)}q + V^{(d)}(1-q)\right)$$

with $q := \frac{\mathrm{e}^{rT}-d}{u-d}$. This formula has the structure of an expectation. In case $0 < q < 1$ (this requires $d < \mathrm{e}^{rT} < u$, a condition guaranteeing absence of arbitrage*), then this q induces a

* What are the arbitrage strategies in case $d \geq \mathrm{e}^{rT}$ or $\mathrm{e}^{rT} \geq u$?

probability Q, and

$$\boxed{V_0 = \mathrm{e}^{-rT}\,\mathsf{E}_{\mathsf{Q}}[V_T]}$$

[Recall that in a discrete probability space with a probability P

$$\mathsf{E}_{\mathsf{P}}[X] = \sum_{i=1}^{n} x_i\ \mathsf{P}(X = x_i)$$

holds, where X is a random variable.] The special probability Q defined above is called **risk-neutral probability**. For S_0 we have

$$\mathsf{E}_{\mathsf{Q}}[S_T] = \underbrace{\frac{\mathrm{e}^{rT} - d}{u - d}}_{=q}\ S_0 u + \underbrace{\frac{u - \mathrm{e}^{rT}}{u - d}}_{=1-q}\ S_0 d = S_0 \mathrm{e}^{rT},$$

or

$$\boxed{S_0 = \mathrm{e}^{-rT}\,\mathsf{E}_{\mathsf{Q}}[S_T]\,.}$$

Summary:

In case the portfolio is risk-free (achieved by the above special value of Δ) and when $0 < q < 1$ with $q = \frac{\mathrm{e}^{rT}-d}{u-d}$, then there is a probability Q, such that

$$V_0 = \mathrm{e}^{-rT}\mathsf{E}_{\mathsf{Q}}[V_T] \quad \text{and}$$
$$S_0 = \mathrm{e}^{-rT}\mathsf{E}_{\mathsf{Q}}[S_T]\,.$$

The quantity Δ is called *Delta*. Later we shall see $\Delta = \frac{\partial V}{\partial S}$ in the time-continuous situation. This is the first and most important example of the "Greeks", others are $\frac{\partial^2 V}{\partial S^2}, \frac{\partial V}{\partial \sigma}, \ldots$

Δ is the key for "*Delta-Hedging*", for minimizing or eliminating the risk of the writer of an option.

Remark: The relation

$$\mathrm{e}^{-rT}\ \mathsf{E}_{\mathsf{Q}}[S_T] = S_0$$

for all T is the *martingale property* of the discounted process $\mathrm{e}^{-rt}S_t$ with respect to the probability Q.

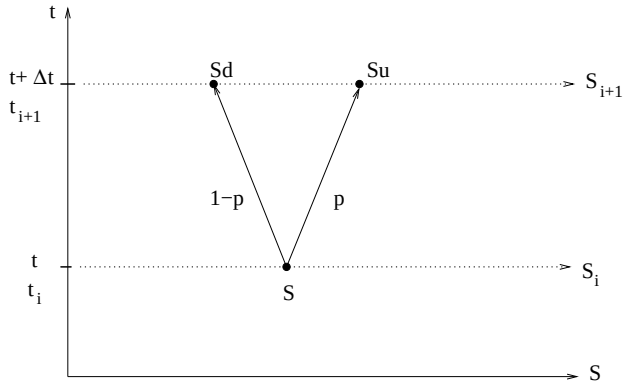
1.3 Binomial Method

For the numerical pricing of options, the continuous time must be discretized. Among the many possible approaches the tree methods have the reputation to be both simple and robust. The simplest version uses a binomial tree. The Black–Scholes model results in the limit when the fineness of the binomial tree goes to zero.

Define an equidistant time discretization as follows:

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

On the S_i-axes we shall define discrete $S_{j,i}$-values.



Assumptions

(Bi1) The market price over one period Δt can only take two values,

$$Su \text{ or } Sd \text{ with } 0 < d < u.$$

(Bi2) Let the probability of an "up" motion be p, $\mathsf{P}(\text{up}) = p$, with $0 < p < 1$.

(Bi3) Expectation and variance equal those of the continuous-time model (for geometric Brownian motion S_t with riskless growth rate r).

(Bi1) and (Bi2) define the framework of a binomial process with probability. The free parameters u, d, p are to be determined such that (Bi1) – (Bi3) hold.

Remarks

1. It turns out that P is the risk-neutral probability Q. Literature on the stochastic background: [Musiela&Rutkowski: Martingale Methods in Financial modeling], [Shreve: Stochastic Calculus for Finance II (Continuous-time models)].
2. In Section 1.5D we shall show for the continuous-time Black–Scholes model

$$\mathsf{E}[S_t] = S_0 \mathrm{e}^{r(t-t_0)}$$
$$\mathsf{E}[S_t^2] = S_0^2 \mathrm{e}^{(2r+\sigma^2)(t-t_0)}$$

   Set S_i for S_0, S_{i+1} for S_t and Δt for $t - t_0$.
3. The expectations are conditional expectations since the initial values $S(t_0)$ or S_i are given.

Conclusion for the step $i \longrightarrow i+1$:

$$\begin{aligned}\mathsf{E}[S(t_{i+1}) \mid S(t_i) = S_i] &= S_i \mathrm{e}^{r\Delta t} \\ \mathrm{Var}[S(t_{i+1}) \mid S(t_i) = S_i] &= S_i^2 \mathrm{e}^{2r\Delta t}(\mathrm{e}^{\sigma^2\Delta t} - 1)\end{aligned}$$

The expectation of the discrete model is

$$\mathsf{E}[S_{i+1}] = p\ S_i u + (1-p)\ S_i d.$$

Equating with the expression of the continuous-time model shows

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

This is the first equation for the three unknowns u, d, p. This gives

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

For $0 < p < 1$ we require

$$d < \mathrm{e}^{r\Delta t} < u\,.$$

This must hold because otherwise arbitrage is possible. (Compare with Section 1.2B to see that p is the q and represents the risk-neutral probability.)
Equating variances leads to

$$\begin{aligned}\mathsf{Var}[S_{i+1}] &= \mathsf{E}[S_{i+1}^2] - (\mathsf{E}[S_{i+1}])^2 \\ &= p\ (S_i u)^2 + (1-p)\ (S_i d)^2 - S_i^2(pu + (1-p)d)^2 \\ &\stackrel{!}{=} S_i^2 \mathrm{e}^{2r\Delta t}(\mathrm{e}^{\sigma^2\Delta t} - 1)\,,\end{aligned}$$

which amounts to

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

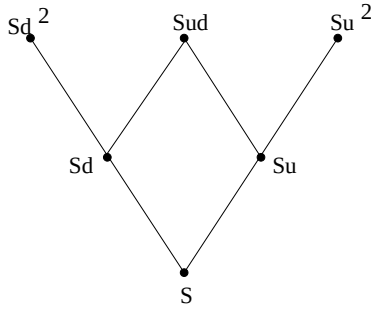
A third equation can be posed arbitrarily. For example, a kind of symmetry is expressed by

$$u \cdot d = 1.$$

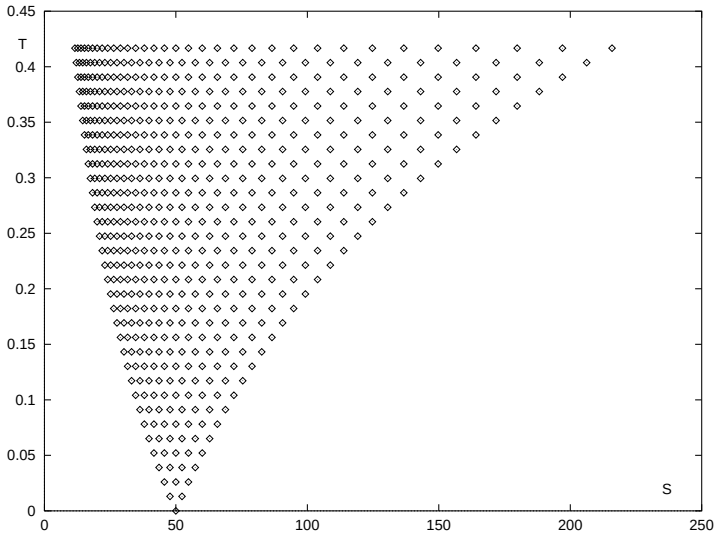
The resulting system of nonlinear equations for u, d, p is

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

This defines the grid of a tree. By the requirement $ud = 1$, this simple tree is rigid in the sense that its parameters u, d, p do not depend on K or S_0. The tree is recombining.



Since $S_{i+1} = \alpha S_i$, $\alpha \in \{u, d\}$ the "branches" of the tree grow exponentially.



The S-values of the grid are

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

Valuation on the Tree

For t_M the value of the option is known from the payoff $\Psi(S) = (S - K)^+$ or $(K - S)^+$:

$$V_{j,M} := \Psi(S_{j,M})$$

By he risk-neutral evaluation principle (Section 1.2B),

$$V_i = \mathrm{e}^{-r\Delta t}\, \mathsf{E}[V_{i+1}],$$

or applied to the tree:

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

This relation establishes a recursion, which starts with $i = M-1$ and prices V at the nodes, until $V_0 := V_{0,0}$.

In case of an **American option**, each node requires a check whether early exercise is reasonable. The holder of the option optimizes her position by comparing the payoff $\Psi(S)$ with the *continuation value*: she chooses the larger value. This requires to modify the above recursion. We denote the continuation value

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

For European options, simply $V_{j,i} := V_{j,i}^{\text{cont}}$. For American options $V_{j,i} := \max\{\Psi(S_{j,i}),\, V_{j,i}^{\text{cont}}\}$, or

$$\begin{aligned} &\text{call:} \quad V_{j,i} := \max\{(S_{j,i}-K)^+, V_{j,i}^{\text{cont}}\} \\ &\text{put:} \quad V_{j,i} := \max\{(K-S_{j,i})^+, V_{j,i}^{\text{cont}}\} \end{aligned}$$

(principle of dynamic programming)

Algorithm (Binomial Method, basic version)

Input: $r,\ \sigma,\ S = S_0,\ T,\ K,$ put or call,
European or American, M
compute: $\Delta t := T/M,\ u,\ d,\ p$ as defined above
$S_{0,0} := S_0$
$S_{j,M} = S_{0,0}u^j d^{M-j},\ j = 0, 1, ..., M$
(for American options in addition $S_{j,i} = S_{0,0}u^j d^{i-j}$
for $0 < i < M$, $j = 0, 1, ..., i$)
$V_{j,M}$ from the payoff
$V_{j,i}$ for $i < M$ by the proper formula
Output: $V_{0,0}$ is approximation for $V(S_0, 0)$

The two different decisions, either *holding* or *exercising* the American-style option, have a geometrical aspect: In the (S,t)-plane the nodes with $V_{j,i}^{\text{cont}} > \Psi(S_{j,i})$ characterize the *continuation area*, and the other nodes are in the *stopping area*. How the early-exercise curve separates the two areas will be discussed in more detail in Chapter 4.

Advantages of the Method

- easy to implement,
- robust, and
- can be adapted to other types of options.

Disadvantages of the Method

- accuracy is rather poor:

  error $O(1/M) = O(\Delta t)$, which is linear convergence. (But the accuracy matches practical requirements.)

- In case V_0 is needed for several values of S, the algorithm must be restarted.

Enhancements

- To avoid oscillations, generalize $ud = 1$ to $ud = \gamma$ and choose γ such that for $t = T$ one node of the tree falls on the strike value K. Then the parameters depend on K and S_0, resulting in a more flexible tree and improved accuracy; extrapolation makes sense.
- Discrete dividend payment at time t_D: Cut the tree at t_D and shift the S-values by $-D$. As result, evaluate the tree at $\tilde{S}_0 := S_0 - D\mathrm{e}^{-rt_D}$. (Illustrations in Topic 1 and 5 in the *Topics for CF*.)
- Sensitivities ("greeks") are calculated by difference quotients.

Problems

In the higher-dimensional case (e.g. basket option with three or more assets) it is not obvious how to generalize the tree.
In the literature the above method is often called Cox-Ross-Rubinstein method (CRR).
Other extensions: trinomial method; "implied grid" for variable $\sigma(S,t)$.

1.4 Stochastic Processes

This section introduces continuous-time models as they are used by Black, Scholes and Merton. Essentially we discuss (geometric) Brownian motion.

History

Brown (1827): studied erratic motion of pollen.

Bachelier (1900): applied Brownian motion to model asset prices.

Einstein (1905): molecular motion

Wiener (1923): mathematical model

since 1940: Itô and others

Definition (Stochastic Process)

A *stochastic process* is a family of random variables X_t for $t \geq 0$ or $0 \leq t \leq T$.

Each sample results in a function X_t called *path* or *trajectory*.

Definition (Wiener process / standard Brownian motion)

W_t (notation also $W(t)$ or W or $\{W_t\}_{t\geq 0}$) has the properties:

(a) W_t is a *continuous* stochastic process

(b) $W_0 = 0$

(c) $W_t \sim \mathcal{N}(0,t)$

(d) All increments $\Delta W_t := W_{t+\Delta t} - W_t$ (Δt arbitrary) on non-overlapping t-intervals are *independent*.

(c) means: W_t is distributed normally with $\mathsf{E}[W_t] = 0$ and $\mathrm{Var}[W_t] = \mathsf{E}[W_t^2] = t$.

Remarks

1) "standard", because it is scalar, driftless, and $W_0 = 0$.
$X_t = a + \mu t + W_t$ with $a, \mu \in \mathbb{R}$ is the general Brownian motion (with drift μ).

2) Consequences (also for $W_0 = a$):

$$\mathsf{E}[W_t - W_s] = 0\ ,\ \mathrm{Var}[W_t - W_s] = t - s \quad \text{for } t > s\,.$$

(show this as exercise)

3) W_t is nowhere differentiable! Motivation:

$$\mathrm{Var}\left[\frac{\Delta W_t}{\Delta t}\right] = \frac{1}{(\Delta t)^2}\mathrm{Var}[\Delta W_t] = \left(\frac{1}{\Delta t}\right)^2 \cdot \Delta t = \frac{1}{\Delta t}$$

tends to ∞ for $\Delta t \to 0$.

4) A Wiener process is self-similar in the sense:

$$W_{\beta t} \stackrel{\mathrm{d}}{=} \sqrt{\beta}\, W_t$$

(both sides obey the same distribution). More general, there are fractal Wiener processes with

$$W_{\beta t} \stackrel{\mathrm{d}}{=} \beta^H W_t\,,$$

for the standard Wiener process $H = \frac{1}{2}$. H is the *Hurst-exponent*. Mandelbrot postulated that finance models should use fractal processes.

Importance

The Wiener process is "driving force" of basic finance models.

Discretization/Computation

So far we have considered W_t for continuous-time models ($t \in \mathbb{R}$). Now we approximate W by a discretization. Take $\Delta t > 0$ as a fixed time increment.

$$t_j := j \cdot \Delta t \quad \Rightarrow \quad W_{j\Delta t} = \sum_{k=1}^{j} (W_{k\Delta t} - W_{(k-1)\Delta t}) = \sum_{k=1}^{j} \Delta W_k$$

The ΔW_k are independent, and by Remark 2 satisfy

$$\mathsf{E}(\Delta W_k) = 0, \quad \mathrm{Var}(\Delta W_k) = \Delta t.$$

In case Z is a random variable with $Z \sim \mathcal{N}(0,1)$ [Chapter 2], then

$$Z\sqrt{\Delta t} \sim \mathcal{N}(0, \Delta t).$$

Hence

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

serves as model for a discretized Wiener process of the ΔW_k.

Algorithm (Simulation of a Wiener process)

start: $t_0 = 0,\ W_0 = 0;$ choose $\Delta t\,.$

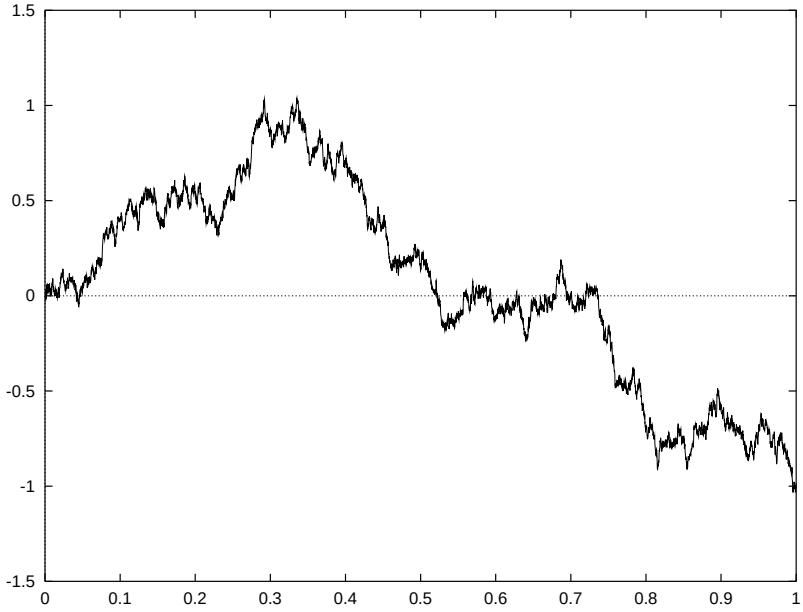
loop $j = 1, 2, \ldots :$

$t_j = t_{j-1} + \Delta t$

draw $Z \sim \mathcal{N}(0,1)$

$W_j = W_{j-1} + Z\sqrt{\Delta t}$

The W_j denotes a realization of W_t at t_j.



a simulation for $\Delta t = 0.0002$

Stochastic Integral

Motivation:

Assume the price of an asset is described by a Wiener process W_t. Let $b(t)$ be the number of assets in the portfolio at time t. For simplicity assume that there are only discrete trading times

$$0 = t_0 < t_1 < \ldots < t_N = T\,.$$

Hence $b(t)$ is piecewise constant:

$$b(t) = b(t_{j-1}) \quad \text{for} \quad t_{j-1} \le t < t_j\,. \qquad (*)$$

The resulting trading gain is

$$\sum_{j=1}^{N} b(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \quad \text{for } 0 \le t \le T.$$

Now we approach the time-continuous case and assume arbitrary trading times. The question is whether the sum converges for $N \to \infty$?

For arbitrary b the integral

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

does not exist as Riemann–Stieltjes integral. Sufficient for its existence would be a finite *first variation* of W_t.

We show: The first variation $\sum_{j=1}^{N} |W_{t_j} - W_{t_{j-1}}|$ is unbounded.

Proof: Clearly

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

for any decomposition of the interval $[0,T]$. Now $\Delta t \to 0$. The second variation is bounded, it converges to a $c \neq 0$ (see the Lemma below). By the continuity of W_t, the first factor of the right-hand side goes to 0, and hence the second factor (the first variation) to ∞.

It remains to investigate what happens with the second variation. The relevant type of convergence is *convergence in the mean*,

$$\lim_{N\to\infty} \mathsf{E}[(X - X_N)^2] = 0\,,$$

$$\text{written as:} \quad X = \operatorname*{l.i.m.}_{N\to\infty} X_N.$$

It remains to show:

Lemma

Denote by $t_0 = t_0^{(N)} < t_1^{(N)} < \ldots < t_N^{(N)} = T$ a sequence of partitions of the interval $t_0 \le t \le T$, with $\delta_N := \max_{j=1}^{N}(t_j^{(N)} - t_{j-1}^{(N)})$. Then:

$$\operatorname*{l.i.m.}_{\delta_N\to 0} \sum_{j=1}^{N} (W_{t_j^{(N)}} - W_{t_{j-1}^{(N)}})^2 = T - t_0$$

Proof: Exercises

Remark: Part of the proof of the lemma comprises the assertions

$$\mathsf{E}[(\Delta W_t)^2 - \Delta t] = 0$$

$$\mathrm{Var}[(\Delta W_t)^2 - \Delta t] = 2 \cdot (\Delta t)^2.$$

In this probabilistic sense the random variable ΔW_t^2 behaves similarly as Δt. Symbolically this is written

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

and will be used for investigations of orders of magnitude.

The construction of an integral for our integrands b

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

is based on $\int_{t_0}^{t} b(s)\mathrm{d}W_s := \sum_{j=1}^{N} b(t_{j-1})(W_{t_j} - W_{t_{j-1}})$ for all step functions b in the sense of $(*)$.

For more general b we take step functions converging to b in the mean. For literature see [Øksendal: Stochastic Differential Equations], [Shreve: Stochastic Calculus].

1.5 Stochastic Differential Equations

A. Integral Equation

Definition (Diffusion model)

The integral equation

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

for a stochstic process X_t is called Itô stochastic differential equation (SDE). Its symbolic notation is

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

Solutions of this stochastic differential equation (that is, of the integral equation) are called *stochastic diffusion*, or *Itô-process*. The term $a(X_t, t)$ is the *drift* term, and $b(X_t, t)$ is the *diffusion*.

Special cases

- The Wiener process is included with $X_t = W_t$, $a = 0$, $b = 1$.
- In the deterministic case $b = 0$ holds, i.e. $\frac{\mathrm{d}X_t}{\mathrm{d}t} = a(X_t, t)$.

Algorithm (analogous as for the Wiener process)

is based on the discrete version

$$\Delta X_t = a(X_t, t)\,\Delta t + b(X_t, t)\,\Delta W_t$$

with ΔW and Δt as in Section 1.4. Let y_j denote an approximation of X_{t_j}.

Start: $t_0,\ y_0 = X_0\,;$ choose $\Delta t\,.$

loop: $j = 0, 1, 2, ...$

$$t_{j+1} = t_j + \Delta t$$
$$\Delta W = Z\sqrt{\Delta t} \quad \text{with} \quad Z \sim \mathcal{N}(0,1)$$
$$y_{j+1} = y_j + a(y_j, t_j)\Delta t + b(y_j, t_j)\Delta W$$

Since $\mathrm{d}W^2 = \mathrm{d}t$, we expect an order of only $\frac{1}{2}$; we come back to this in Chapter 3.

B. Application to the Stock Market

Model (GBM = geometric Brownian motion)

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

This is an Itô-stochastic SDE with $a = \mu S_t$ and $b = \sigma S_t$. This SDE is *linear* as long as μ and σ do not depend on S_t. For Black and Scholes μ and σ are constant.

(This fills the gap GBM in Assumption 3 in Section 1.2 in the market model.)

μ is interpreted as growth rate, and σ as volatility. The relative change is described by

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

The classic theory of Black, Scholes and Merton (and a significant part of this chapter) assumes a GBM with constant μ, σ.

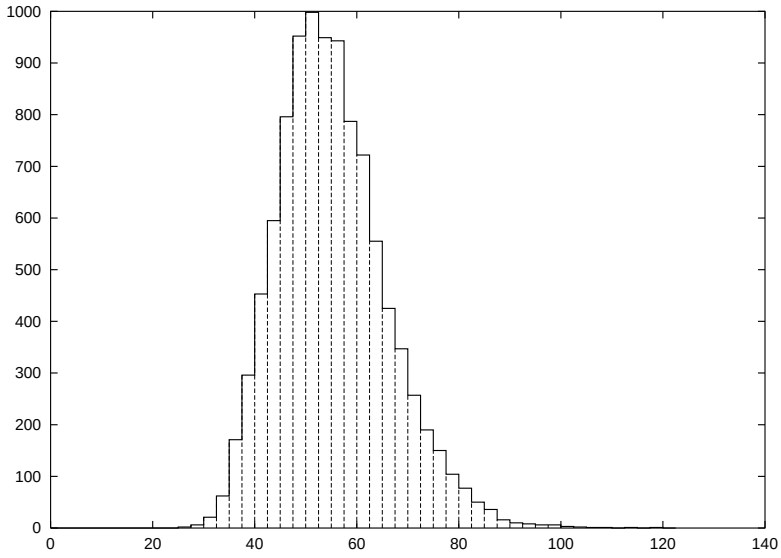
(Bachelier's model was

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

here the price S_t can become negative.)

Recommendation

Implement the algorithm (with Z from Chapter 2), and integrate the GBM for a chosen set of parameters (for instance $\mu = 0.1$, $\sigma = 0.2$) 10000 times until $t = 1$. Then distribute the obtained values S_1 in subintervals, and count the values. This yields a histogram reflecting a lognormal distribution (see figure).



Consequence:

From

$$\frac{\Delta S}{S} = \mu\Delta t + \sigma\,\Delta W$$

we conclude for the distribution of the $\frac{\Delta S}{S}$:

1) distributed normally
2) $\mathsf{E}[\frac{\Delta S}{S}] = \mu\Delta t$
3) $\mathrm{Var}[\frac{\Delta S}{S}] = \sigma^2\Delta t$

together: $\frac{\Delta S}{S} \sim \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$

This offers a way to calculate volatilities σ empirically: For a sequence of trading days collect the data $\frac{\Delta S}{S}$, call them R_i *(returns)*, where R_{i+1} and R_i are measured at time distance Δt. Assuming that GBM is appropriate to describe the returns, σ is obtained as

$$\sigma = \frac{1}{\sqrt{\Delta t}} * \text{ standard deviation of the } R_i\,.$$

This specific value of σ, based on data of the past, is called *historic volatility* (for the *implied* volatility see the Exercises.)

S under GBM can be approximated by the above algorithm as long as $\Delta t > 0$ is small enough, and $S > 0$.

Other models

GBM is continuous, and its density has thin tails, which often fails to describe real asset prices observed in the market. Therefore also other stochastic processes are used, as jump processes, or processes with stochastic volatility. In the following, we stick to the Itô-SDEs, that is to continuous processes driven by Wiener process.

Mean reversion (often used for interest rate models)

Here R denotes an average level of interest rate. Let us investigate the SDE

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

for a stochastic process r_t. That is,

$$a(r_t, t) = \alpha(R - r_t) \quad \textit{mean reversion} \text{ drift}$$
$$b(r_t, t) = \sigma^{\mathrm{r}}\, r_t^{\beta}$$

with suitable parameters $R, \alpha, \sigma^{\mathrm{r}}, \beta$ (obtained by calibration). This has the effect on the drift:

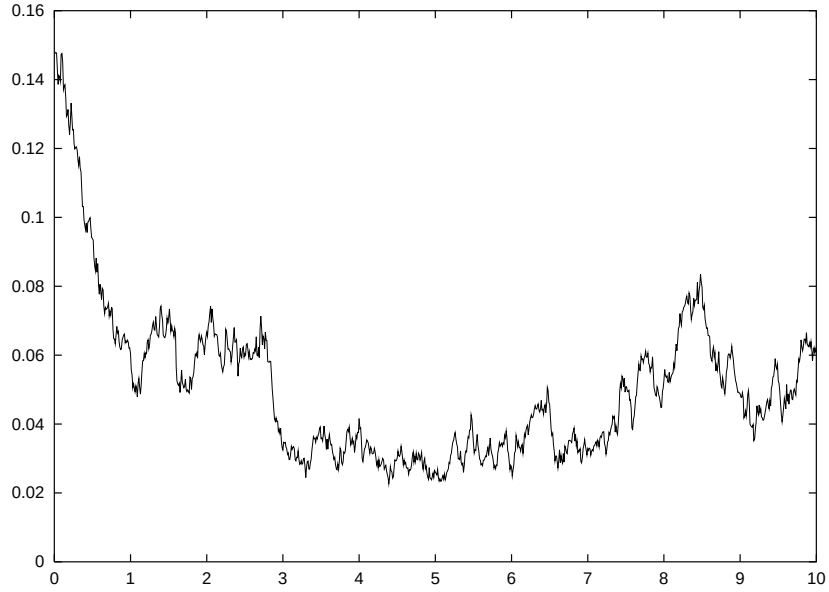
$r_t < R \quad \Rightarrow$ positive growth rate

$r_t > R \quad \Rightarrow$ decay

This effect is superseded by the stochastic fluctuations, but essentially the mean reversion takes care that the order of magnitude of r_t stays close to R, or reverts to R. The parameter α controls the intensity of the reversion.

For $\beta = \frac{1}{2}$, i.e. $b(r_t, t) = \sigma^{\mathrm{r}}\sqrt{r_t}$, the model is called CIR model (Cox-Ingersoll-Ross model).

A simulation r_t of the Cox-Ingersoll-Ross model for $R = 0.05$, $\alpha = 1$, $\beta = 0.5$, $r_0 = 0.15$, $\sigma^{\mathrm{r}} = 0.1$, $\Delta t = 0.01$



The next extension is to:

Vector-valued processes

Assume $W_t = (W_t^{(1)}, \ldots, W_t^{(m)})$ is a m-dimensional Brownian motion. Define for $i = 1, ..., n$

$$X_t^{(i)} = X_{t_0}^{(i)} + \int_{t_0}^{t} a_i(X_s, s)\,\mathrm{d}s + \sum_{k=1}^{m} \int_{t_0}^{t} b_{i,k}(X_s, s)\,\mathrm{d}W_s^{(k)},$$

with vectors

$$X_t = \begin{pmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(n)} \end{pmatrix} \quad , \quad a(X_s, s) = \begin{pmatrix} a_1(X_s^{(1)}, \ldots, X_s^{(n)}, s) \\ \vdots \\ a_n(X_s^{(1)}, \ldots, X_s^{(n)}, s) \end{pmatrix}$$

and matrix

$$\big(\big(b_{i,k}\big)\big)_{i=1,\ldots,n}^{k=1,\ldots,m}$$

which involves the covariances of the vector process.

Example 1 Heston's model

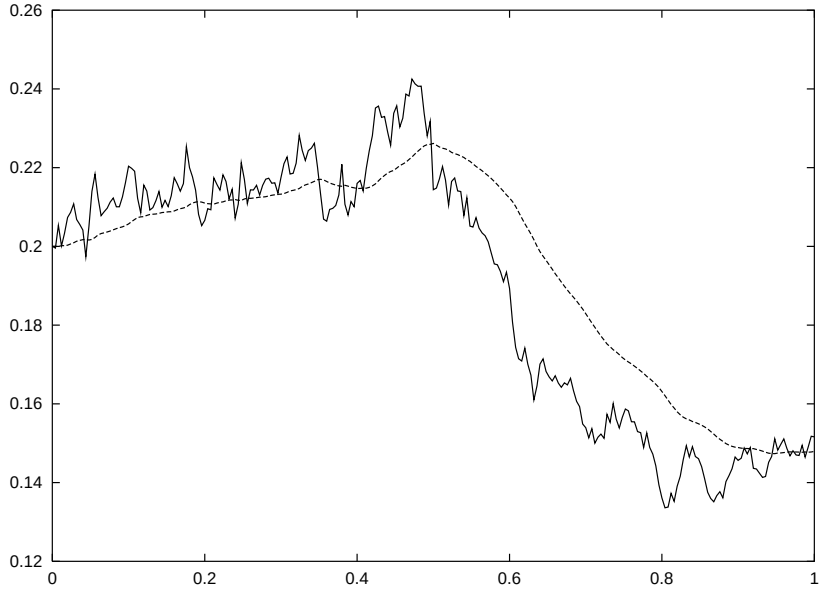
$$\begin{aligned} \mathrm{d}S_t &= \mu S_t\,\mathrm{d}t + \sqrt{v_t} S_t\,\mathrm{d}W^{(1)} \\ \mathrm{d}v_t &= \kappa(\theta - v_t)\,\mathrm{d}t + \sigma^{\mathsf{vola}}\sqrt{v_t}\,\mathrm{d}W^{(2)} \end{aligned}$$

The **stochastic volatility** $\sqrt{v_t}$ is defined via a mean reversion for the variance v_t. This model (with $n = 2$ and $m = 2$) involves parameters $\kappa, \theta, \sigma^{\mathsf{vola}}$, the correlation ρ

between $W^{(1)}$ and $W^{(2)}$, an initial value v_0 and a growth rate μ which may be given by a risk-free valuation concept. Altogether, about five parameters must be calibrated. Heston's model is used frequently.

Example 2 volatility tandem

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



$\alpha = 0.3$, $\beta = 10$; dashed: ζ

Hint: *local* volatility means

$$\sigma = \sigma(t, S_t).$$

C. Itô Lemma

Motivation (deterministic case)

Suppose $x(t)$ is a function, and $y(t) := g(x(t), t)$. The chain rule implies

$$\frac{\mathrm{d}}{\mathrm{d}t}g = \frac{\partial g}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial g}{\partial t}.$$

With $\mathrm{d}x = a(x(t), t)\mathrm{d}t$ this can be written

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

Lemma (Itô)

Assume X_t is an Itô process following $\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\,\mathrm{d}W_t$ and $g(x, t) \in \mathrm{C}^2$. Then $Y_t := g(X_t, t)$ solves the SDE

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

That is, Y_t is an Itô process with the same Wiener process as the input process X.

Sketch of a proof:

$$\left.\begin{array}{c} t \to t+\Delta t \\ X \to X+\Delta X \end{array}\right\} \to \; g(X+\Delta X, t+\Delta t) = Y + \Delta Y$$

Taylor expansion of g gives ΔY as follows:

$$\Delta Y = \frac{\partial g}{\partial X} \cdot \Delta X + \frac{\partial g}{\partial t}\Delta t + \text{terms quadratic in } \Delta t, \Delta X$$

Substitute

$$\begin{aligned} \Delta X &= a\,\Delta t + b\,\Delta W \\ (\Delta X)^2 &= a^2\,\Delta t^2 + b^2 \underbrace{\Delta W^2}_{=O(\Delta t)} + 2ab\,\Delta t\,\Delta W \end{aligned}$$

and order the terms according to powers of Δt, ΔW to obtain

$$\Delta Y = \left(\frac{\partial g}{\partial X}\,a + \frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}\,b^2\right)\Delta t + b\,\frac{\partial g}{\partial X}\,\Delta W + \text{t.h.o.}$$

Similar as in Section 1.4, ΔW can be written as sum, and *convergence in the mean* is applied. See [Øksendal].

D. Application to the GBM model

Assume the GBM model

$$\mathrm{d}S = \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W$$

with μ and σ constant, i.e. $X = S$, $a = \mu S$, $b = \sigma S$.

1) Let $V(S,t)$ be smooth ($\in C^2$)

$$\Rightarrow \quad \mathrm{d}V = \left(\frac{\partial V}{\partial S}\mu S + \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right)\mathrm{d}t + \frac{\partial V}{\partial S}\sigma S\,\mathrm{d}W$$

This is the basic SDE which leads to the PDE of Black and Scholes for the value function $V(S,t)$ of a European standard option.

2) $Y_t := \log(S_t)$, i.e. $g(x) = \log x$

$$\begin{aligned} &\Rightarrow \quad \frac{\partial g}{\partial x} = \frac{1}{x} \quad \text{and} \quad \frac{\partial^2 g}{\partial x^2} = -\frac{1}{x^2} \\ &\Rightarrow \quad \mathrm{d}\left(\log S_t\right) = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t \end{aligned}$$

Hence the log-prices $Y_t = \log S_t$ satisfy a simple SDE, with the elementary solution:

$$Y_t = Y_{t_0} + \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W_t - W_{t_0})$$

$$\Rightarrow \quad \log S_t - \log S_{t_0} = \log \frac{S_t}{S_{t_0}} = \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W_t - W_{t_0})$$

$$\Rightarrow \quad S_t = S_{t_0} \cdot \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W_t - W_{t_0})\right]$$

For $t_0 = 0$ and $W_{t_0} = W_0 = 0$, this results in

$$\boxed{S_t = S_0 \exp\left[\left(\mu - \tfrac{\sigma^2}{2}\right) t + \sigma W_t\right]}$$

In summary, S_t is exponential function of a Brownian motion with drift.

Implications for $t_0 = 0$:

a) $\log S_t$ is distributed normally

b) $\mathsf{E}[\log S_t] = \mathsf{E}[\log S_0] + (\mu - \frac{\sigma^2}{2})t + 0 = \log S_0 + (\mu - \frac{\sigma^2}{2})t$

c) $\mathsf{Var}[\log S_t] = \mathsf{Var}[\sigma W_t] = \sigma^2 t$

summarizing a) – c) means

$$\log \frac{S_t}{S_0} \sim \mathcal{N}\left((\mu - \frac{\sigma^2}{2})t\,,\, \sigma^2 t\right)$$

d) This leads to the density function of $Y = \log S$

$$\hat{f}(Y) = \hat{f}(\log S_t) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left[-\frac{(\log(S_t/S_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}\right].$$

And what is the density of S_t? The probabilities of S and Y are the same and hence also the distribution integrals. We apply the transformation theorem (Section 2.2B) for $Y := \log S$ and have the integrands

$$\hat{f}(Y)\,\mathrm{d}Y = \underbrace{\hat{f}(\log S)\frac{1}{S}}_{f(S_t)}\,\mathrm{d}S.$$

Consequently, the density f of the distribution of the asset price S_t is

$$f(S_t, t;\, S_0, \mu, \sigma) := \frac{1}{S_t\sigma\sqrt{2\pi t}} \exp\left[-\frac{(\log(S_t/S_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}\right].$$

This is the density f_{GBM} of the *lognormal distribution*. It describes the probability of the transition $(S_0, 0) \longrightarrow (S_t, t)$ under GBM.

e) Now the last gap in the derivation of the binomial method can be closed: There the continuous model refers to our GBM. As an exercise, realize

$$\mathsf{E}(S) = \int_0^\infty S f(\ldots)\,\mathrm{d}S = S_0\,\mathrm{e}^{\mu(t-t_0)}$$

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

1.6 Risk-Neutral Valuation

(This section *sketches* basic ideas and concepts. For a thorough treatment we recommend literature on Stochastic Finance such as [Musiela&Rutkowski: Martingale Methods in Financial modeling])

Recall (from the one-period)

$$V_0 = \mathrm{e}^{-rT}\mathsf{E}_\mathsf{Q}[\Psi(S_T)]$$

where Q is the artificial probability of Section 1.2 and $\Psi(S_T)$ denotes the payoff.

For the model with continuous time formally the same relation holds. But Q and E_Q are different. It turns out that the density of Q is given by $f(S_t, t;\, S_0, r, \sigma)$, with μ replaced by r. Hence the relation

$$V_0 = \mathrm{e}^{-rT} \int_0^\infty \Psi(S_T) \cdot f(S_T, T;\, S_0, r, \sigma)\, \mathrm{d}S_T$$

holds for the GBM-based continuous model. In the following we outline the arguments that lead to this integral.

Fundamental Theorem of Asset Pricing

The market model is free of arbitrage if and only if there is a probability Q such that the discounted asset prices $\mathrm{e}^{-rt}S_t$ are martingales with respect to Q.

Probability space

The same sample space and σ-algebra $(\Omega, \mathcal{F})$ underlying a Wiener process are not specified. The chosen probability P completes $(\Omega, \mathcal{F})$ to the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. The independence of the increments ΔW of the Wiener process depend on P. A process W can be a Wiener process with respect to P, but is no Wiener process with respect to another probability $\hat{\mathsf{P}}$

Martingale

A martingale M_t is a stochastic process with

$$\mathsf{E}[M_t \mid \mathcal{F}_s] = M_s \quad \text{for all } t, s \text{ with } s \le t\,,$$

where $\mathcal{F}_s$ is a filtration, i.e. a family of σ-algebras with $\mathcal{F}_s \subseteq \mathcal{F}_t\ \forall s \le t$. A filtration serves as model for the amount of information in a market.

$\mathsf{E}[M_t \mid \mathcal{F}_s]$ is a *conditional expectation*. It can be regarded as expectation of M_t conditional on the amount of information available until time instant s.

M_t martingale means that M_s at time s is the best possible forecast for $t \ge s$.

Martingale with respect to a probability Q: $\mathsf{E}_\mathsf{Q}[M_t \mid \mathcal{F}_s] = M_s$ for all t, s with $s \le t$.

Examples of martingales

1) any Wiener process
2) $W_t^2 - t$ for any Wiener process W.
3) A necessary criterion for martingales is the absence of drift.

Essentially, drift-free processes are martingales.

Market Price of Risk

$$\begin{aligned} \mathrm{d}S &= \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W \\ &= rS\,\mathrm{d}t + (\mu - r)S\,\mathrm{d}t + \sigma S\,\mathrm{d}W \\ &= rS\,\mathrm{d}t + \sigma S\left[\frac{\mu - r}{\sigma}\,\mathrm{d}t + \,\mathrm{d}W\right] \end{aligned}$$

The investor expects $\mu > r$ as a compensation for the risk, which is represented by σ. $\mu - r$ is the *excess return.*

$$\begin{aligned} \gamma := \frac{\mu - r}{\sigma} &= \text{ “market price of risk”} \\ &= \text{ compensation rate relative to the risk} \end{aligned}$$

Hence

$$\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,[\gamma\,\mathrm{d}t + \mathrm{d}W] \qquad (*)$$

For the probability P the term in brackets represnts a drifted Brownian motion and no (standard) Wiener process.

Girsanov's Theorem

Suppose W is Wiener process with respect to $(\Omega, \mathcal{F}, \mathsf{P})$. In case γ satisfies certain requirements, there is a probability Q such that

$$W_t^\gamma := W_t + \int_0^t \gamma\,\mathrm{d}s$$

is a (standard) Wiener process under Q.

(probability theory: Q results from the Theorem of Radon-Nikodym. Q and P are equivalent. For constant γ the requirements of Girsanov are fulfilled.)

Application

Substitute $\mathrm{d}W^\gamma = \mathrm{d}W + \gamma \mathrm{d}t$ in $(*)$ gives

$$\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,\mathrm{d}W^\gamma.$$

This is a change of drift from μ to r; σ remains unchanged. The path of S_t under the probability Q is defined by the density $f(\ldots, r, \sigma)$. The transition from $f(\ldots, \mu, \sigma)$ to $f(\ldots, r, \sigma)$ amounts to adjusting the probability from P to Q. The discounted $\mathrm{e}^{-rt}S_t$ is drift-free under Q and Martingale. Q is called “risk-neutral” probability.

Trading Strategy

Let X_t be a stochastic vector process of market prices, and b_t denotes the vector with the numbers of shares held in the portfolio. Hence $b_t^{\mathrm{tr}} X_t$ is the wealth process of the portfolio.

Example

$$X_t := \begin{pmatrix} S_t \\ B_t \end{pmatrix},$$

where S_t is the market price of the asset underlying an option, and B_t is the value of a risk-free bond.

Notation: V_t is the random variable of the value of an European option.
Assumptions:

(1) There is a strategy b_t *replicating* the payoff of the option at time T,

$$b_T^{\mathrm{tr}} X_T = \text{ Payoff } .$$

b_t must be $\mathcal{F}_t$-measurable for all t. (That is, the trader cannot see the future. Note that the value of the payoff is a random variable.)

(2) The portfolio is closed, no money is inserted or withdrawn. This is the *self-financing property* defined as

$$\mathrm{d}(b_t^{\mathrm{tr}} X_t) = b_t^{\mathrm{tr}} \mathrm{d}X_t \, .$$

(3) The market is free of arbitrage.

(1), (2), (3) $\Rightarrow V_t = b_t^{\mathrm{tr}} X_t$ for $0 \le t \le T$ (otherwise there would exist arbitrage)

We consider a European option and a discounting process Y_t with the property that $Y_t X_t$ is martingale. Then one can show that also $Y_t b_t^{\mathrm{tr}} X_t$ is martingale (both with respect to Q).

Implications for European options for $t \le T$

$$\begin{aligned} Y_t V_t = Y_t b_t^{\mathrm{tr}} X_t &= \mathsf{E}_{\mathsf{Q}}[Y_T b_T^{\mathrm{tr}} X_T \mid \mathcal{F}_t] \quad \text{(martingale)} \\ &= \mathsf{E}_{\mathsf{Q}}[Y_T \cdot \text{Payoff} \mid \mathcal{F}_t] \quad \text{(replication)} \end{aligned}$$

When the payoff is a function Ψ of S_T (vanilla-option under GBM), then

$$= \mathsf{E}_{\mathsf{Q}}[Y_T \cdot \Psi(S_T)]$$

(because $W_T - W_t$ is independent of $\mathcal{F}_t$). Discounting with $Y_t = \mathrm{e}^{-rt}$ implies specifically for $t = 0$

$$1 \cdot V_0 = \mathsf{E}_{\mathsf{Q}}[\mathrm{e}^{-rT} \cdot \Psi(S_T)]$$

and hence

$$V_0 = \mathrm{e}^{-rT} \int_0^\infty \Psi(S_T) \cdot f(S_T, T;\, S_0, r, \sigma) \, \mathrm{d}S_T \, .$$

This is called *risk-neutral valuation.*

Literature on Stochastic Finance: [Elliot & Kopp: Mathematics of Financial Markets], [Korn: Option Pricing and Portfolio Optimization], [Musiela & Rutkowski: Martingale Methods in Financial modeling], [Shreve: Stochastic Calculus for Finance].

Outlook

We so far have investigated continuous processes S_t driven by W_t. To compensate for occasional drastic changes in the price of underlying, one resorts to models with stochastic volatility, or to jump processes.

Supplements

The "**Greeks**" mean the sensitivities of $V(S,t;\,\sigma,r)$ and are defined as

$$\text{Delta} = \frac{\partial V}{\partial S}, \quad \text{gamma} = \frac{\partial^2 V}{\partial S^2}, \quad \text{theta} = \frac{\partial V}{\partial t}, \quad \text{vega} = \frac{\partial V}{\partial \sigma}, \quad \text{rho} = \frac{\partial V}{\partial r}$$

Black–Scholes Formula

For a European call the analytic solution of the Black–Scholes equation is

$$\begin{aligned} d_1 &:= \frac{\log \frac{S}{K} + \left(r - \delta + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &:= d_1 - \sigma\sqrt{T-t} = \frac{\log \frac{S}{K} + \left(r - \delta - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ V_{\rm C}(S,t) &= S\mathrm{e}^{-\delta(T-t)}F(d_1) - K\mathrm{e}^{-r(T-t)}F(d_2), \end{aligned}$$

where F denotes the standard normal cumulative distribution (compare Exercises), and δ is a continuous dividend yield. The value $V_{\rm P}(S,t)$ of a put is obtained by applying the put-call parity

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

from which

$$V_{\rm P} = -S\mathrm{e}^{-\delta(T-t)}F(-d_1) + K\mathrm{e}^{-r(T-t)}F(-d_2)$$

follows.

2. Computation of Random Numbers

Definition (sample from a distribution)

A sequence of numbers is called *sample* from a distribution function F, if the numbers are independent realizations of a random variable with distribution F.

Examples

If F is the uniform distribution on the interval $[0,1]$, then we call the samples from F *uniform deviates*. Notation: $\sim \mathcal{U}[0,1]$.

If F is the standard normal distribution, then we call the samples from F *standard normal deviates*. Notation: $\sim \mathcal{N}(0,1)$.

The basis of random number generation is to draw numbers $\sim \mathcal{U}[0,1]$.

2.1 Uniform Deviates

A. Linear Congruential Generators

Choose $a, b, M \in \mathbb{N}$, $a \neq 0$, $a, b < M$, and define for $N_0 \in \mathbb{N}$ (*"seed"*) a sequence of numbers by

Algorithm (linear congruential generator)

choose N_0 .
For $i = 1, 2, ...$ calculate
$N_i = (aN_{i-1} + b) \bmod M$

Define $U_i \in [0,1)$ by

$$U_i = \frac{N_i}{M} .$$

The numbers U_i are used as uniform deviates.

Obvious **Properties**

(a) $N_i \in \{0, 1, ..., M-1\}$

(b) The sequence of N_i is periodic with a period $p \leq M$.
(*because* there are not $M+1$ distinct numbers N_i. Hence two out of $\{N_0, ..., N_M\}$ must be equal, $N_i = N_{i+p}$ with $p \leq M$. p-periodicity follows.)

Literature: [D. Knuth: The Art of Computer Programming, Volume 2]

The above numbers U_i are no real random numbers, but are deterministically defined and reproducible. We call such numbers *pseudo random*. In this chapter, we omit the modifier "pseudo" because it is clear from the context. The aim is to find parameters M, a, b such that the numbers U_i are good substitutes of real random numbers.

Example

$$M = 244944, \quad a = 1597, \quad b = 51749$$

Useful parameters a, b, M are in [Press et al.: Numerical Recipes].

Question: What are "good" random numbers?

A practical (and hypothetical) answer: The numbers should pass "all" tests.

First requirement: The period p must be large, hence M as large as possible. For example, in a binary computer with mantissa length l, one aims at $M \approx 2^l$. Suitable a, b can be derived with methods from number theory. [Knuth].

Second requirement: The numbers must be distributed as intended (density f, expectation μ, variance σ^2). Check this by **statistical tests** as follows: First apply the algorithm to produce a large number of U_i-values. Then

(a) Calculate the mean $\hat{\mu}$ and the variance $\hat{s}^2$ of the sample. Check $\hat{\mu} \approx \mu$ and $\hat{s}^2 \approx \sigma^2$.

(b) Test for correlations of the U_i with previous U_{i-j}. For example, correlation could mean that small values of U are likely to be followed again by small values. In this case the generator would be of low quality.

(c) Estimate the density function $\hat{f}$ of the sample, and check for $\hat{f} \approx f$. A prototypical test is as follows: Divide the unit interval $[0, 1]$ into equidistant subintervals

$$k\Delta U \le U < (k+1)\Delta U\,,$$

where ΔU denotes the length of the subintervals. (For other distributions choose an interval that contains all sample points U_i, and the subintervals will be defined accordingly.) When altogether j samples are calculated, let j_k be the number of samples that fall into the kth subinterval. The probability that the kth subinterval is hit is $\frac{j_k}{j}$. This should approximate

$$\int_{k\Delta U}^{(k+1)\Delta U} f(x)\,\mathrm{d}x \quad (f = 1 \text{ for the uniform distribution})\,.$$

This integral is

$$\Delta U f(\bar{U})\,,$$

with $\bar{U}$ in the kth subinterval. Hence a good generator should satisfy

$$\Delta U \hat{f}(\bar{U}) = \frac{j_k}{j} \stackrel{!}{=} \Delta U f(\bar{U})\,,$$

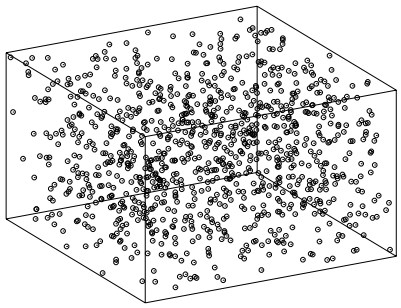
at least for small ΔU. The empirical density on the kth subinterval is

$$\hat{f} = \frac{j_k}{j\Delta U}\,.$$

Third requirement: The **lattice structure** should be OK. To check this, arrange vectors out of m consecutive numbers:

$$(U_i, U_{i+1}, \ldots, U_{i+m-1})$$

For $U \sim \mathcal{U}[0,1]$, these points should fill the m-dimensional unit-cube as uniformly as possible. The sequences of points/vectors lie on $(m-1)$-dimensional hyperplanes. Trivial case: M parallel planes through $U = \frac{i}{M}, \quad i = 0, \ldots, M-1$ (any of the m components).



A bad situation occurs when all points fall on only a *few* planes. Then the gaps between the planes without any points would be wide. This leads to analyze the lattice structure of the random points. The focus is on the smallest number of planes, on which all points in $[0,1)^m$ "land."

Analysis for $m = 2$: In this planar case, the hyperplanes in (U_{i-1}, U_i)-space are straight lines $z_0 U_{i-1} + z_1 U_i = \lambda$, for parameters z_0, z_1, λ. From

$$\begin{aligned} N_i &= (aN_{i-1} + b) \bmod M \\ &= aN_{i-1} + b - kM \quad \text{for } \; kM \le aN_{i-1} + b < (k+1)M \end{aligned}$$

conclude for arbitrary numbers z_0, z_1

$$\begin{aligned} z_0 N_{i-1} + z_1 N_i &= z_0 N_{i-1} + z_1(aN_{i-1} + b - kM) \\ &= N_{i-1}(z_0 + az_1) + z_1 b - z_1 kM \\ &= M \underbrace{\left(N_{i-1}\frac{z_0 + az_1}{M} - z_1 k\right)}_{=:c=c(i)} + z_1 b \end{aligned}$$

Dividing by M leads to

$$z_0 U_{i-1} + z_1 U_i = c + z_1 b M^{-1}\,,$$

a straight line in the (U_{i-1}, U_i)-plane. For fixed z_0, z_1 this defines a family of parallel lines/"planes," parameterized by c.

Question: Is there a family of such lines (planes) defined by a pair (z_0, z_1), such that only few lines (planes) cut the unit-cube? The minimal number of parallel hyperplanes holding all points is the worst case.

For analyzing the number of planes, the cardinality of the c's matters. To find the worst case with a small set of c's, assume $z_1, z_0 \in \mathbb{Z}$ and $z_0 + az_1 \bmod M = 0$. Then the parameter c is integer, and

$$c = z_0 U_{i-1} + z_1 U_i - z_1 b M^{-1} \in \mathbb{Z}\,.$$

($z_1 b M^{-1}$ is a constant parallel shift not affecting the number of planes.) How many of such c's exist? For $0 \le U < 1$ obtain a range for the c's by a maximal set $I_c \subset \mathbb{Z}$, such that

$$c \in I_c \quad \Rightarrow \quad \text{the line touches or cuts the unit-cube .}$$

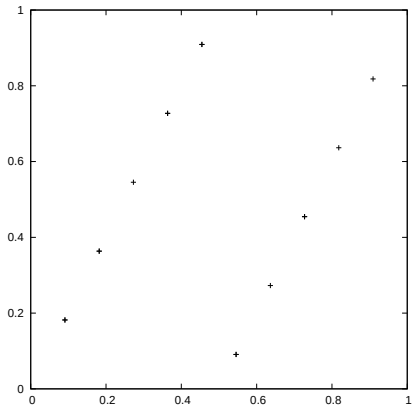
The cardinality of the set I_c gives a clue on the distance between the parallel lines (planes). It is unfavorable when the set consists of only a few elements.

Academic Example $N_i = 2N_{i-1} \mod 11$ (i.e. $a = 2,\ b = 0,\ M = 11$)

The pair $(z_0, z_1) = (-2, 1)$ solves $z_0 + az_1 = 0 \mod M$. Hence

$$-2U_{i-1} + U_i = c\,.$$

$0 \le U < 1$ implies $-2 < c < 1$. In view of $c \in \mathbb{Z}$, the only parameters are $c = -1$ and $c = 0$. For this choice of (z_0, z_1) all 10 points in $[0,1)^2$ fall on only two straight lines. (0 does not occur for $N_0 \ne 11k$, $k \in \mathbb{Z}$.)



(U_{i-1}, U_i)-plane.
10 points for $N_0 \ne 11k$

Example $N_i = (1229N_{i-1} + 1) \mod 2048$

The condition $z_0 + az_1 = 0 \mod M$

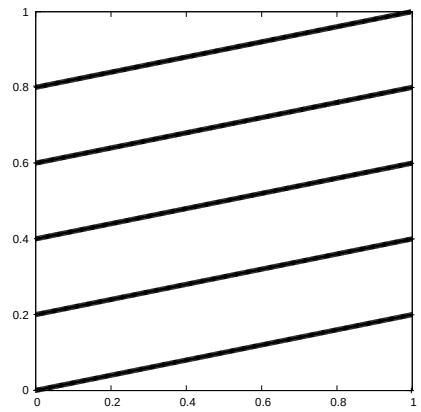
$$\frac{z_0 + 1229 z_1}{2048} \in \mathbb{Z}$$

is satisfied by $z_0 = -1,\ z_1 = 5$, because

$$-1 + 1229 \cdot 5 = 6144 = 3 \cdot 2048\,.$$

$c = -U_{i-1} + 5U_i - \frac{5}{2048}$ implies $-1 - \frac{5}{2048} < c < 5 - \frac{5}{2048}$. This shows that there are only six values of the c's, $c \in \{-1, 0, 1, 2, 3, 4\}$, and all points in $[0,1)^2$ fall on six straight lines.

The U_i-distance between two neighboring lines is $\frac{1}{z_1} = \frac{1}{5}$.



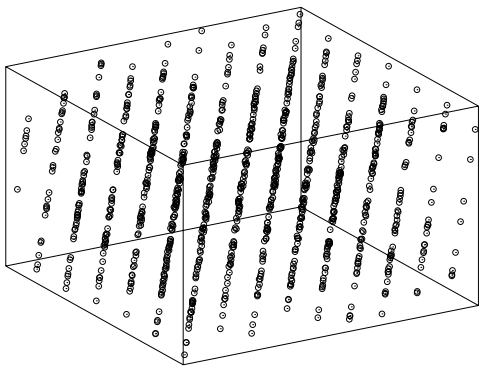
(U_{i-1}, U_i)-plane.
In this figure,
the discrete points
are not separated.
The sixth line
consists of one point.

For the above examples, the (U_{i-1}, U_i)-points are obviously not equidistributed. The next example is much better for $m = 2$. But equidistrution for $m = 2$ does not imply equidistribution for larger m.

Example (RANDU)

$$N_i = aN_{i-1} \bmod M, \quad \text{with } a = 2^{16} + 3, \ M = 2^{31}$$

For $m = 2$ experiments show that the dots (U_{i-1}, U_i) are nicely equidistributed in the square. But inspection for $m = 3$ reveals that the random points in the cube $[0, 1)^3$ fall on only 15 planes.



Analysis for larger m is analogous. Illustration in *Topic 14*.

B. Fibonacci Generators

There are other classes of random-number generators, for example, the **Fibonacci generators**. A prototype of such generators is defined by

$$N_{i+1} := N_{i-\nu} - N_{i-\mu} \bmod M$$

for suitable μ, ν (also with "+" or with more terms). Literature: [Knuth]

Example

$$U_i := U_{i-17} - U_{i-5},$$
$$\text{in case } U_i < 0 \text{ set } U_i := U_i + 1.0$$

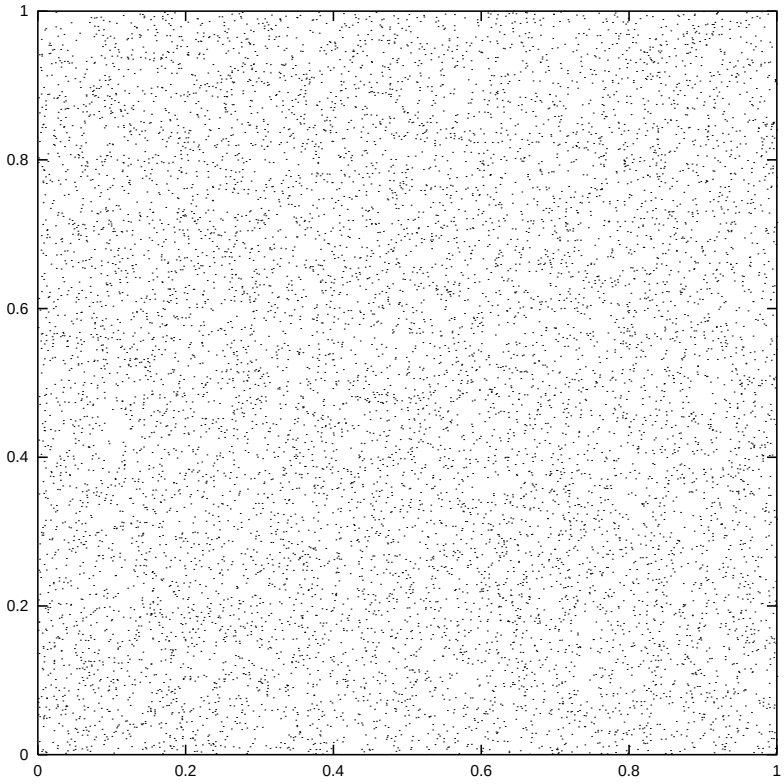
(simple example with reasonable features, but there are correlations.)

Algorithm (loop of a simple Fibonacci generator)

Repeat: $\zeta := U_i - U_j$
in case $\zeta < 0$, set $\zeta := \zeta + 1$
$U_i := \zeta$
$i := i - 1$
$j := j - 1$
in case $i = 0$, set $i := 17$
in case $j = 0$, set $j := 17$

Initialization: Set $i = 17,\ j = 5$, and calculate $U_1, ..., U_{17}$ with a congruential generator with, for example, $M = 714025,\ a = 1366,\ b = 150889$.

A professional generator to calculate uniform random numbers is the "Mersenne Twister" by Matsumoto, Nishimura, in: ACM Transactions on Modelling and Computer Simulations **8** (1998), p.3-30. This generator has excellent features, with a huge period, and is equidistributed also for high dimensions m.



10000 random numbers (U_{i-1}, U_i), calculated with a Fibonacci Generator

2.2 Random Numbers from Other Distributions

The generation of all kind of deviates is based on uniform deviates. For the calculation of random numbers from a given distribution we can apply several methods, namely, inversion, transformations, and rejection methods.

A. Inversion

Let $F(x) := \mathsf{P}(X \le x)$ be a distribution function, for a random variable X, and P is the corresponding probability.

Theorem (inversion)

Suppose $U \sim \mathcal{U}[0,1]$ and let F be a continuous strictly increasing distribution function. Then $X := F^{-1}(U)$ is a sample from F.

Proof:

$U \sim \mathcal{U}[0,1]$ means $\mathsf{P}(U \le \xi) = \xi$ for $0 \le \xi \le 1$. Hence

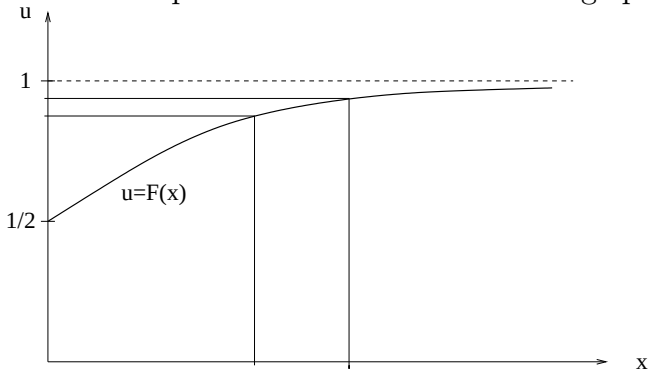
$$\mathsf{P}(F^{-1}(U) \le x) = \mathsf{P}(U \le F(x)) = F(x).$$

Application

Calculate $u \sim \mathcal{U}[0,1]$ and evaluate $F^{-1}(u)$. These numbers have the desired distribution. Mostly inversion is done numerically because F^{-1} in general is not known analytically.

There are two variants:

(a) $F(x) = u$ is a nonlinear equation for x, which can be solved iteratively with standard methods of numerical analysis (e.g. Newton method). For the normal distribution (Figure), the iteration requires tricky termination criteria, because for $u \approx 0,\ u \approx 1$ small perturbations in u lead to large perturbations in x.



(b) Construct an approximating function G such that $G(u) \approx F^{-1}(u)$. Then only $x = G(u)$ needs to be evaluated. The construction of G must observe the asymptotic behavior, which amounts to the poles of G. For the standard normal distribution the symmetry w.r.t. $(x,u) = (0, \frac{1}{2})$ can be exploited and only the pole for $u = 1$ needs to be observed. This can be done with a rational function $G(u)$, with a denominator having a zero at $u = 1$.

B. Transformation

We begin with the scalar case: Let X be a random variable. What is the distribution of a transformed $h(X)$?

Theorem (scalar transformation)

Suppose X is a random variable with density function f and distribution function F. Further assume

$$h : S \to B$$

with $S, B \subseteq \mathbb{R}$, where S is the support of f, and let h be strictly monotonic.

(a) $Y := h(X)$ is random variable with distribution function

$$\begin{array}{ll} F(h^{-1}(y)) & \text{for increasing } h \\ 1 - F(h^{-1}(y)) & \text{for decreasing } h \end{array}$$

(b) If h^{-1} is absolutely continuous, then for almost all y the density of $h(X)$ is

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| .$$

Proof: (write also F^X for F)

(a) $F^Y(y) := \mathsf{P}(h(X) \le y) =$

(in case h is increasing:)

$= \mathsf{P}(X \le h^{-1}(y)) = F^X(h^{-1}(y))$

(in case h is decreasing:)

$= \mathsf{P}(X \ge h^{-1}(y)) = 1 - \mathsf{P}(X < h^{-1}(y)) = 1 - F^X(h^{-1}(y))$

(b) For absolutely continuous h^{-1} the density of $Y = h(X)$ is equal to the derivative of the distribution function almost everywhere. Evaluation of $\frac{\mathrm{d}F(h^{-1}(y))}{\mathrm{d}y}$ with the chain rule implies the assertion; distinguish between increasing and decreasing h.

Application

Start with $X \sim \mathcal{U}[0,1]$ and the density of the uniform distribution,

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

i.e. $S = [0,1]$. Random numbers Y with prescribed target density $g(y)$ are to be calculated. Hence we require a transformation h such that

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = 1 \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| \stackrel{!}{=} g(y) .$$

Then $h(X)$ is distributed as intended.

Example (exponential distribution)

The exponential distribution with parameter $\lambda > 0$ has the density

$$g(y) = \begin{cases} \lambda \mathrm{e}^{-\lambda y} & \text{for } y \geq 0 \\ 0 & \text{for } y < 0. \end{cases}$$

B consists of the non-negative real numbers. As transformation $[0,1] \to B$ we choose the monotonic decreasing function

$$y = h(x) := -\frac{1}{\lambda} \log x$$

with inverse $h^{-1}(y) = \mathrm{e}^{-\lambda y}$ for $y \geq 0$. Since

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = 1 \cdot \left| (-\lambda)\mathrm{e}^{-\lambda y} \right| = \lambda \mathrm{e}^{-\lambda y} \ = g(y)\,,$$

$h(X)$ is distributed exponentially for $X \sim \mathcal{U}[0,1]$.

Application

Calculate $U_1, U_2, ... \sim \mathcal{U}[0,1]$. Then

$$-\frac{1}{\lambda}\log(U_1), \quad -\frac{1}{\lambda}\log(U_2), \quad ... \text{ are distributed exponentially.}$$

(Hint: The distances between jump times of Poisson processes are distributed exponentially.)

Attempt with the normal distribution: Search for h such that

$$1 \cdot \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right).$$

This is a differential equation for h^{-1} without analytic solution. In this situation the multi-dimensional version of the transformation helps.

Theorem (transformation in $\mathbb{R}^n$)

Suppose X is a random variable in $\mathbb{R}^n$ with density $f(x) > 0$ on the support S. Let the transformation $h : S \to B$, $S, B \subseteq \mathbb{R}^n$ be invertible and the inverse continuously differentiable on B. Then $Y := h(X)$ has the density

$$f(h^{-1}(y)) \left| \frac{\partial(x_1, ..., x_n)}{\partial(y_1, ..., y_n)} \right|, \quad y \in B, \tag{2.7}$$

where $\frac{\partial(x_1,...,x_n)}{\partial(y_1,...,y_n)}$ denotes the determinant of the Jacobian matrix of $h^{-1}(y)$.

Proof: see Theorem 4.2 in [L. Devroye: Non-Uniform Random Variate Generation (1986)]

In Section 2.3 the two-dimensional version will be applied to calculate normal variates.

C. Acceptance-Rejection Method

This method is based on the following facts: Let f be a density function on $S \subset \mathbb{R}$ and $\mathcal{A}_f$ the area between the x-axis and the graph of f. Assume two random variables U and X independent of each other with $U \sim \mathcal{U}[0,1]$ and X distributed with density f. Then the points

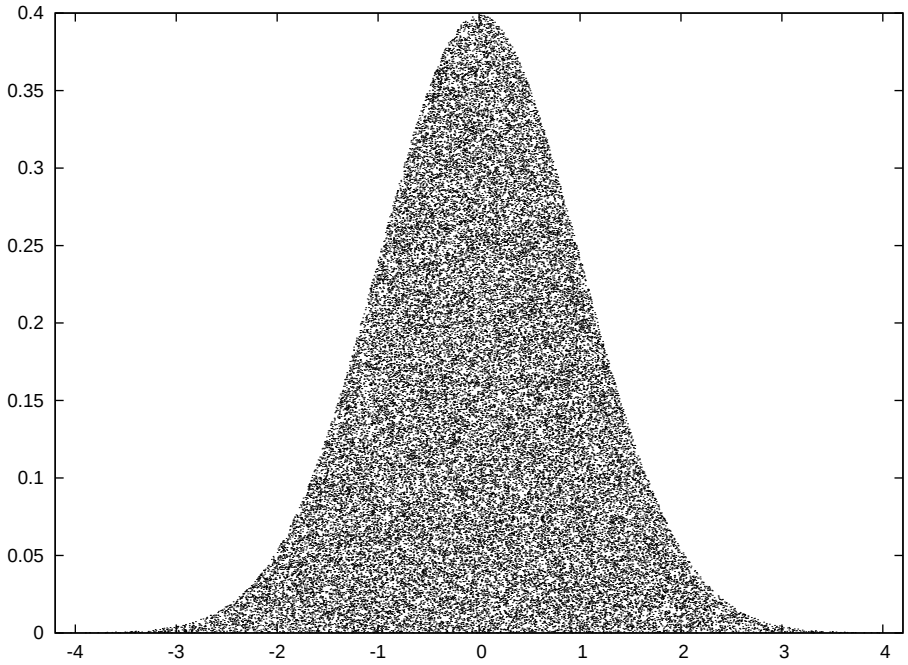
$$(x, y) := (X,\, U \cdot f(X))$$

are uniformly distributed on $\mathcal{A}_f$ (and vice versa). In the Figure this is illustrated for the normal distribution. If one cuts off a piece of the area $\mathcal{A}_f$, then the remaining points are still distributed uniformly. This is exploited by rejection methods.

Let g be another density on S, and assume for a constant $c \geq 1$

$$f(x) \leq c\, g(x) \quad \text{for all } x \in S\,.$$

The function cg is major to f, and the set $\mathcal{A}_f$ is subset of the area $\mathcal{A}_{cg}$ underneath the graph of cg. A rejection algorithm assumes that g-distributed x-samples can be calculated easily. Then the points $(x, ucg(x))$ are distributed uniformly on $\mathcal{A}_{cg}$. The aim is to calculate f-distributed random numbers. Cutting off the part of $\mathcal{A}_{cg}$ above $\mathcal{A}_f$ means to reject points with $ucg(x) > f(x)$. The x-coordinates of the remaining points with $ucg(x) \leq f(x)$ are accepted and are distributed as desired.



50000 points
$(X,\, U \cdot f(X))$
$X \sim \mathcal{N}(0,1)$
$U \sim \mathcal{U}[0,1]$
f density of X

Repeat:
  $x :=$ random number distributed according to g,
  $u :=$ random number $\sim \mathcal{U}[0,1]$ independent of x
until $u\, c\, g(x) \leq f(x)$
return: x

Example (as exercise): Laplace-density $g(x) := \frac{1}{2}\exp(-|x|)$, f density of the standard normal distribution. What is c?*

2.3 Normal Deviates

This section applies the transformation theorem in $\mathbb{R}^2$ to the calculation of normally distributed random numbers, and sketches the ziggurat algorithm. (Alternative methods are provided by inversion methods.)

A. Method of Box and Muller

$S := [0,1]^2$, X uniformly distributed on S, density $f = 1$ on S. Transformation h:

$$\begin{cases} y_1 = \sqrt{-2\log x_1}\,\cos 2\pi x_2 =: h_1(x_1,x_2) \\ y_2 = \sqrt{-2\log x_1}\,\sin 2\pi x_2 =: h_2(x_1,x_2) \end{cases}$$

inverse h^{-1}:

$$\begin{cases} x_1 = \exp\left\{-\frac{1}{2}(y_1^2+y_2^2)\right\} \\ x_2 = \dfrac{1}{2\pi}\arctan\dfrac{y_2}{y_1} \end{cases}$$

For this transformation the determinant is

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \\ = -\frac{1}{2\pi}\exp\left\{-\tfrac{1}{2}(y_1^2+y_2^2)\right\}\,.$$

Its absolute value is the density of the two-dimensional normal distribution. Since

$$\left|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| = \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}y_1^2\right)\right]\cdot\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}y_2^2\right)\right]\,,$$

the two-dimensional density is the product of the one-dimensional densities of the standard normal distribution. As a consequence, the two components y_1, y_2 of the vector Y are independent.

Application

When the two components x_1, x_2 are distributed $\sim \mathcal{U}[0,1]$, then the transformation provides two independent $y_1, y_2 \sim \mathcal{N}(0,1)$.

Algorithm (Box-Muller)

(1) generate $U_1 \sim \mathcal{U}[0,1]$ and $U_2 \sim \mathcal{U}[0,1]$.
(2) $\theta := 2\pi U_2, \quad \rho := \sqrt{-2\log U_1}$
(3) $Z_1 := \rho\cos\theta$ is $\sim \mathcal{N}(0,1)$
(same as $Z_2 := \rho\sin\theta$).

* colored in Topic 3 of the *Topics for CF*

B. Variant of Marsaglia

Prepare the input x_1, x_2 for the Box–Muller transformation such that trigonometric functions are avoided. From $U \sim \mathcal{U}[0,1]$ obtain $V := 2U - 1 \sim \mathcal{U}[-1,1]$. Two such numbers V_1, V_2 define a point in $\mathbb{R}^2$. Define the disk

$$\mathcal{D} := \{(V_1, V_2) \ : \ V_1^2 + V_2^2 < 1\}.$$

Accept only those pairs (U_1, U_2) such that $(V_1, V_2) \in \mathcal{D}$. These accepted points are uniformly distributed on $\mathcal{D}$. Transformation to (radius)^2 and normalized angle:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} V_1^2 + V_2^2 \\ \frac{1}{2\pi} \arg(V_1, V_2) \end{pmatrix}.$$

These (x_1, x_2) are distributed uniformly in S ($\longrightarrow$ exercise) and serve as input for Box&Muller. The advantage is:

$$\cos(2\pi x_2) = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$
$$\sin(2\pi x_2) = \frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$

Algorithm (polar method)

(1) *Repeat:* generate $U_1, U_2 \sim \mathcal{U}[0,1]$;
$V_1 := 2U_1 - 1$, $V_2 := 2U_2 - 1$;
until $w := V_1^2 + V_2^2 < 1$.

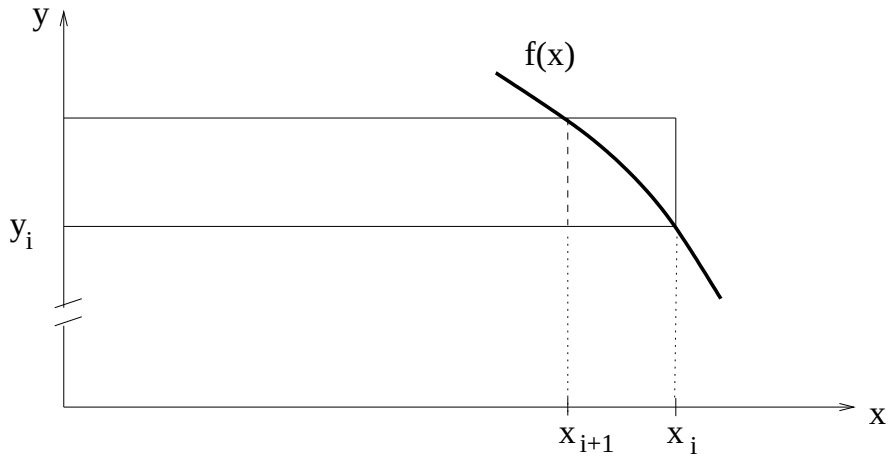
(2) $Z_1 := V_1 \sqrt{-2\log(w)/w}$ is $\sim \mathcal{N}(0,1)$
(as well as $Z_2 := V_2 \sqrt{-2\log(w)/w}$).

The probability of acceptance ($w < 1$) is the ratio of the areas $\frac{\pi}{4} \approx 0.785...$. That is, 21% of all draws (U_1, U_2) are rejected. But these costs are compensated by the saving of trigonometric functions, and Marsaglia's polar method is more efficient than standard Box&Muller.

C. Ziggurat Algorithm

A most efficient algorithm for the generation of normal deviates is the ziggurat algorithm, which is a rejection method.

Essentiall, g is a step function $\geq$ the Gaussian density f. Construction with N horizontal layers of $N-1$ rectangles *with the same area*, and one bottom segment with the same area, which is no rectangle but infinite because of the tail of f. The figure illustrates schematically a situation for $x \geq 0$, where the rectangle consists of two portions (for each i with $0 < i < N-1$), which make an extremely efficient test for acceptance possible (random choice of the layer i; uniformly distributed test point).



2.4 Correlated Normal Random Variates

The aim is the generation of a normal random vector $X = (X_1, \ldots, X_n)$ with prescribed

$$\mu = \mathsf{E}X = (\mathsf{E}X_1, \ldots, \mathsf{E}X_n),$$

covariance matrix with elements

$$\Sigma_{ij} = (\mathsf{Cov}X)_{ij} := \mathsf{E}\left((X_i - \mu_i)(X_j - \mu_j)\right); \quad \sigma_i^2 = \Sigma_{ii}$$

and correlations

$$\rho_{ij} := \frac{\Sigma_{ij}}{\sigma_i \sigma_j}\,.$$

For the following assume that Σ is symmetric and positive definite.

Recall: The density function $f(x_1, \ldots, x_n)$ of $\mathcal{N}(\mu, \Sigma)$ is

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^{tr}\Sigma^{-1}(x-\mu)\right\}.$$

Assume $Z \sim \mathcal{N}(0, I)$, where z is a realization of Z and I the unit matrix. We apply the linear transformation $x = Az$, $A \in \mathbb{R}^{n\times n}$ nonsingular. The transformation theorem yields with $X = h(Z) := AZ$ the density of X as

$$\begin{aligned} f(A^{-1}x)\,|\det(A^{-1})| &= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(A^{-1}x)^{tr}(A^{-1}x)\right\} \frac{1}{|\det(A)|} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(A)|} \exp\left\{-\frac{1}{2}x^{tr}(AA^{tr})^{-1}x\right\} \end{aligned}$$

for arbitrary nonsingular matrices A. In case AA^{tr} is a factorization of Σ, $\Sigma = AA^{tr}$, and hence $|\det A| = (\det \Sigma)^{1/2}$, we conclude:

$$AZ \sim \mathcal{N}(0, \Sigma)\,.$$

This implies

$$\mu + AZ \sim \mathcal{N}(\mu, \Sigma)\,.$$

Example: Choose the Cholesky decomposition of Σ.
Alternative decomposition out of a principal component analysis of Σ.

Algorithm (correlated normal deviates)

(1) Decompose Σ into $AA^{tr} = \Sigma$
(2) Draw $Z \sim \mathcal{N}(0, I)$ componentwise
with $Z_i \sim \mathcal{N}(0,1)$ for $i = 1, ..., n$, for example,
with Marsaglia's polar method
(3) $\mu + AZ$ is distributed $\sim \mathcal{N}(\mu, \Sigma)$

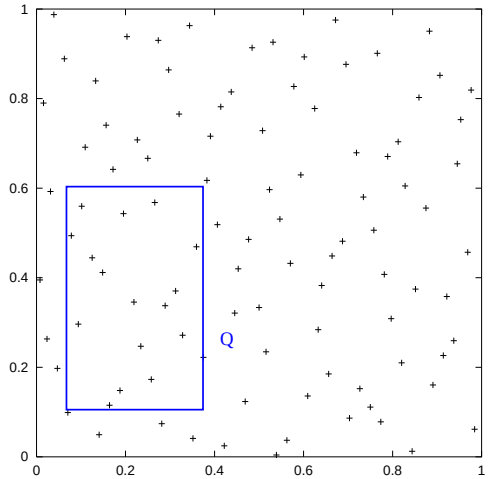
Example: If $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is required, the solution is

$$\begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2 \end{pmatrix}\,.$$

2.5 Sequences of Numbers with Low Discrepancy

The aim is to construct points distributed similarly as random numbers, but avoid clustering or holes. In order to characterize **equidistributedness**, take any box (hyperrectangle) in $[0,1]^m$, $m \geq 1$. It would be desirable if for all Q

$$\frac{\#\text{ of the } x_i \in Q}{\#\text{ all points in } [0,1]^m} \approx \frac{\text{vol}(Q)}{\text{vol}([0,1]^m)}$$



For $m = 2$ the figure illustrates this idea in the unit square $[0,1]^2$.

Definition (discrepancy)

The discrepancy of a set $\{x_1, \ldots, x_N\}$ of N points with $x_i \in [0,1]^m$ is

$$D_N := \sup_Q \left| \frac{\#\text{ of the } x_i \in Q}{N} - \text{ vol}(Q) \right| .$$

We wish to find sequences of points, whose discrepancy D_N for $N \to \infty$ tends to zero "quickly." To assess the decay we compare with the sequence

$$\frac{1}{\sqrt{N}} ,$$

which characterizes the probabilistic error of Monte Carlo methods. For true random points the discrepancy has a similar order of magnitude, namely,

$$\sqrt{\frac{\log\log N}{N}} .$$

Definition (sequence of low discrepancy)

A sequence of points $x_1, \ldots, x_N, \ldots \in [0,1]^m$ is called low-discrepancy sequence if there is a constant C_m such that for all N

$$D_N \leq C_m \frac{(\log N)^m}{N} .$$

Comment

The denominator in $\frac{1}{N}$ stands for relatively rapid decay of D_N with the number of points N, rapid as compared with the $\frac{1}{\sqrt{N}}$ of Monte Carlo.

But we have to observe the numerator $(\log N)^m$. Since $\log N$ grows only modestly, for low dimension m the decay of D_N is much faster than the decay of the probabilistic Monte Carlo error.

Table: different convergence rates to zero

N	$\frac{1}{\sqrt{N}}$	$\sqrt{\frac{\log\log N}{N}}$	$\frac{\log N}{N}$	$\frac{(\log N)^2}{N}$	$\frac{(\log N)^3}{N}$
10^1	.31622777	.28879620	.23025851	.53018981	1.22080716
10^2	.10000000	.12357911	.04605170	.21207592	.97664572
10^3	.03162278	.04396186	.00690776	.04771708	.32961793
10^4	.01000000	.01490076	.00092103	.00848304	.07813166
10^5	.00316228	.00494315	.00011513	.00132547	.01526009
10^6	.00100000	.00162043	.00001382	.00019087	.00263694
10^7	.00031623	.00052725	.00000161	.00002598	.00041874
10^8	.00010000	.00017069	.00000018	.00000339	.00006251
10^9	.00003162	.00005506	.00000002	.00000043	.00000890

Do sequences of low discrepancy exist?

Example: $(m = 1)$ **Van der Corput sequence**

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \ldots$$

Let us study its construction by means of the example $x_6 = \frac{3}{8}$. The binary representation of the index 6 is 110. This is radix-inverted: .011, which gives $\frac{3}{8}$.

Definition (radical-inverse function)

For $i = 1, 2, ...$ let

$$i = \sum_{k=0}^{j} d_k b^k$$

be the expansion in base b (integer ≥ 2), with $d_k \in \{0, 1, \ldots, b-1\}$. The radical-inverse function is defined by

$$\phi_b(i) := \sum_{k=0}^{j} d_k b^{-k-1} .$$

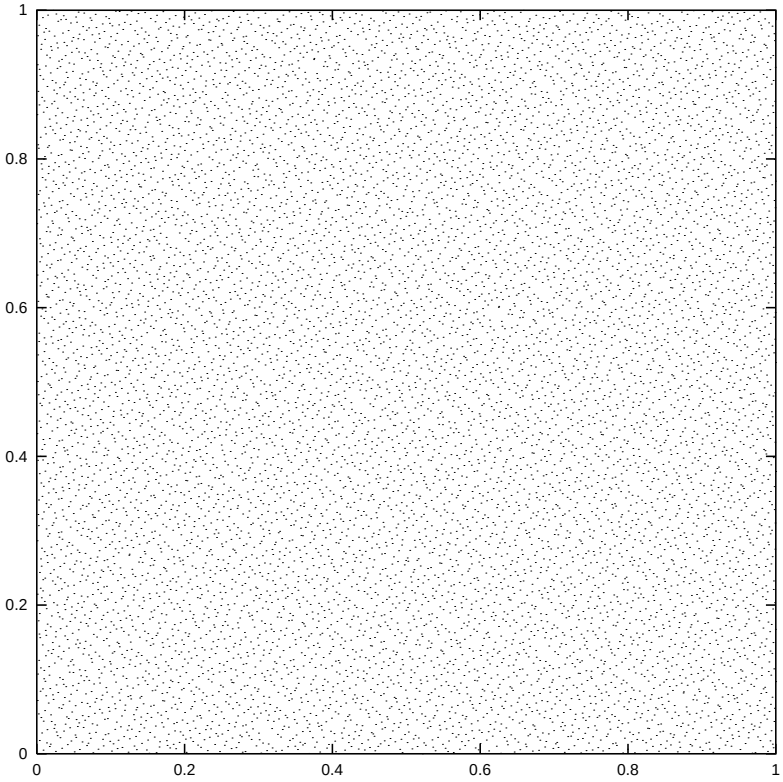
A one-dimensional example is the Van der Corput sequence: $x_i := \phi_2(i)$.

Definition (Halton sequence)

Let $p_1, \ldots, p_m$ be pairwise prime integers. The *Halton sequence* is defined as the sequence of vectors

$$x_i := (\phi_{p_1}(i), \ldots, \phi_{p_m}(i)) , \quad i = 1, 2, ...$$

The Halton sequence is of low discrepancy with $C_2 = 0.2602$ for $m = 2$ and easy to generate.



The figure shows the first 10000 Halton points with $m = 2$ and $p_1 = 2$, $p_2 = 3$.

Other sequences of low discrepancy:

- Faure sequence
- Sobol sequence
- Niederreiter sequence
- Halton "leaped": For large m the Halton sequence suffers from correlation. This can be cured by taking

$$x_i := (\phi_{p_1}(li), \ldots, \phi_{p_m}(li)), \quad i = 1, 2, ...$$

  for suitable prime l different from the p_k, for example, $l = 409$.

The deterministic sequences of low discrepancy are called **quasi-random numbers**. (They are not random!)

Literature on quasi-random numbers: [H. Niederreiter: Random Number Generation and Quasi-Monte Carlo Methods (1992)]

3. Monte Carlo Methods

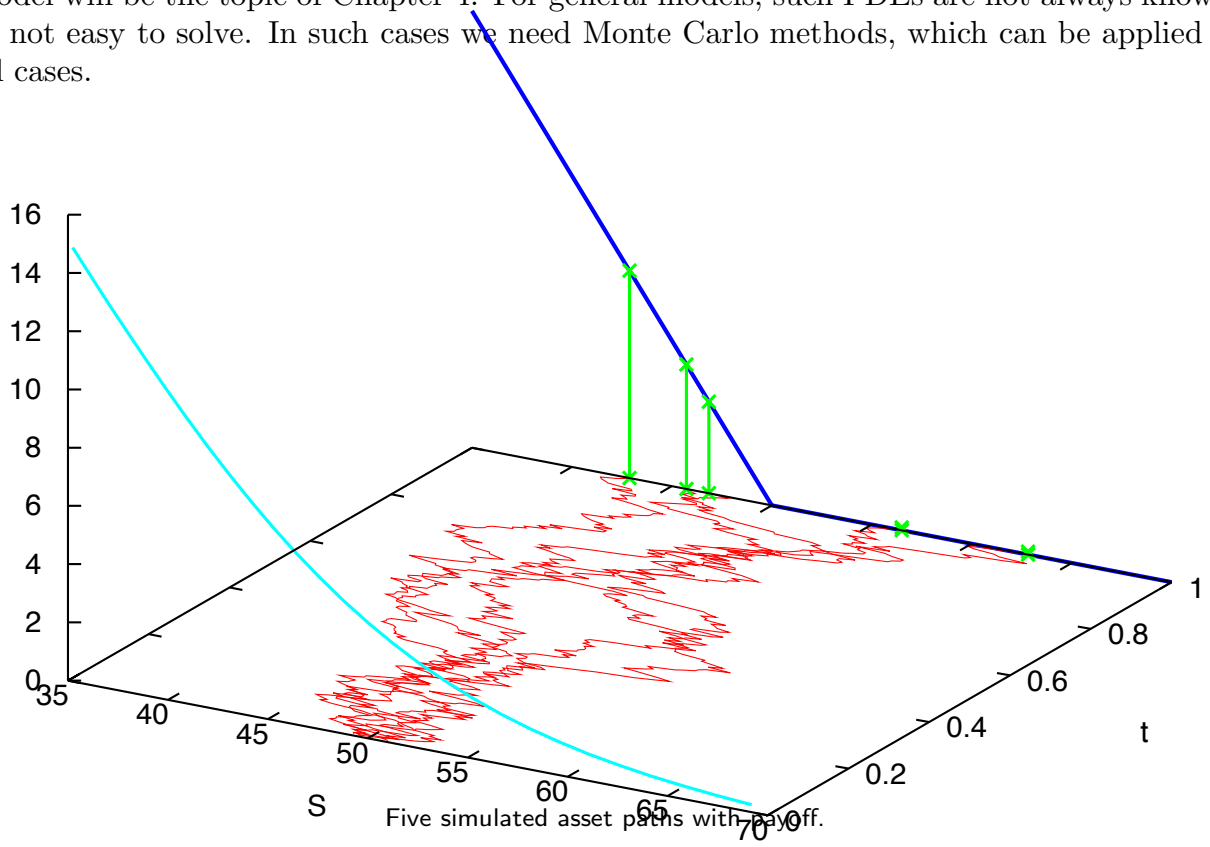
In Chapter 1 we introduced the formula of risk-neutral valuation of European options,

$$V(S_0, 0) = \mathrm{e}^{-rT} \mathsf{E}_{\mathsf{Q}} \left[\Psi(S_T) \mid S_0 \right] ,$$

where $\Psi(S_T)$ denotes the payoff. In the Black–Scholes model, specifically, this is

$$V(S_0, 0) = \mathrm{e}^{-rT} \int_0^\infty \Psi(S_T) \cdot f_{\mathrm{GBM}}(S_T, T; S_0, r, \sigma) \, \mathrm{d}S_T \, . \qquad \text{(Int)}$$

(For the transition density f_{GBM} see Section 1.5D.) The resulting PDE of the Black–Scholes model will be the topic of Chapter 4. For general models, such PDEs are not always known, or not easy to solve. In such cases we need Monte Carlo methods, which can be applied in all cases.



Five simulated asset paths with payoff.

There are two approaches to calculate the above integral:

1) The integral (Int) is approximated using numerical quadrature.
2) One applies Monte Carlo simulation. That is, one draws random numbers that match the underlying risk-neutral probability, and calculates many paths of asset prices S_t. This is the bulk of the work. To complete, compute the mean of the payoff values, and discount.

In this chapter we confine ourselves to the second approach.

Notations (from Chapter 1)

A scalar SDE driven by a Wiener process is described by

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

We discretize time t with a grid

$$\ldots < t_{j-1} < t_j < t_{j+1} < \ldots\ ,$$

with equidistant step h or $\Delta t = t_{j+1} - t_j$. Let y_j denote an approximation of X_{t_j}, where $y_0 := X_0$.

Example: **Euler discretization**

$$\begin{aligned} y_{j+1} &= y_j + a(y_j,t_j)\,\Delta t + b(y_j,t_j)\,\Delta W_j\,, \\ t_j &= j\Delta t\,, \\ \Delta W_j &= W_{t_{j+1}} - W_{t_j} = Z\sqrt{\Delta t} \\ &\text{with } Z \sim \mathcal{N}(0,1)\,. \end{aligned} \tag{Euler}$$

3.1 Approximation Error

Definition

For a given path of the Wiener processes W_t we call a solution X_t of (SDE) a **strong solution**. In case the Wiener process is free, X_t or (X_t, W_t) is called **weak solution**.

For strong solutions the numerical discretization is based on the same W_t as the SDE. This enables to investigate the pathwise difference $X_t - y_t$ for convergence behavior for $h \to 0$.

Notation: We write y_t^h for a numerically (with step length h) calculated approximation y at t, in particular, for $t = T$.

Definition (absolute error)

For a strong solution X_t of (SDE) and an approximation y_t^h the absolute error at $t = T$ is defined as

$$\varepsilon(h) := \mathsf{E}\big[\ |X_T - y_T^h|\ \big]\ .$$

For a GBM, where the analytic solution X_t is known, $\varepsilon(h)$ can be obtained easily:

Set in (SDE) $a(X_t,t) = \alpha X_t$ and $b(X_t,t) = \beta X_t$. Then the solution (see Section 1.5D) for given W_T is

$$X_T = X_0 \exp\left[\left(\alpha - \frac{\beta^2}{2}\right)T + \beta W_T\right]\ .$$

The expectation $\varepsilon(h)$ can be estimated as mean of a large number of evaluations of $|X_T - y_T^h|$. For Euler's method this empirical investigation reveals the error behavior

$$\varepsilon(h) = O(h^{\frac{1}{2}}) \,,$$

which is a low accuracy compared to the deterministic case $O(h)$. The result is plausible because ΔW is of the order $O(\sqrt{h})$ (in probability), compare Section 1.4.

Definition (strong convergence)

y_T^h converges strongly to X_T with order $\gamma > 0$, if

$$\varepsilon(h) = \mathsf{E}\big[\ |X_T - y_T^h|\ \big] = O(h^\gamma) \,.$$

y_T^h converges strongly if

$$\lim_{h\to 0} \mathsf{E}\big[\ |X_T - y_T^h|\ \big] = 0 \,.$$

Example

When a and b satisfy global Lipschitz conditions and bounded growth conditions, then the Euler discretization converges strongly with order $\gamma = \frac{1}{2}$.

Note that for several important SDE models (such as CIR, Heston) the global Lipschitz conditions do not hold. Then modifications of the standard Euler may be necessary, also to guarantee $S_t \geq 0$.

How about weak solutions?

In many practical situations the individual paths of X_t are not of interest. Instead, the focus may be on *moments* of X_T. In particular, we would like to know $\mathsf{E}[X_T]$ or $\mathsf{Var}[X_T]$, rather than samples of X_T. For options, the interest is on $\mathsf{E}[\Psi(X_T)]$.

Definition (weak convergence)

y_T^h converges weakly to X_T with respect to a function g with order $\beta > 0$, if

$$\mathsf{E}[g(X_T)] - \mathsf{E}\big[g(y_T^h)\big] = O(h^\beta) \,,$$

and converges weakly with order β, if this holds for all polynomials g.

Example

When a and b are four times continuously differentiable, the Euler method is weakly convergent with order $\beta = 1$.

Importance of g

In case the convergence order β holds for all polynomials, the convergence of all moments follows.

Proof (for the first two moments):

(a) For $g(x) := x$

$$\mathsf{E}[X_T] - \mathsf{E}[y_T^h] = O(h^\beta)$$

holds, viz, convergence of the mean.

(b) If in addition the convergence order holds for $g(x) := x^2$, then (writing $y := y_T^h$ and $X := X_T$)

$$\begin{aligned}\left|\mathsf{Var}[X_T] - \mathsf{Var}[y_T^h]\right| &= \left|\mathsf{E}[X^2] - \mathsf{E}[y^2] - (\mathsf{E}[X])^2 + (\mathsf{E}[y])^2\right| \\ &\le \underbrace{|\mathsf{E}[X^2] - \mathsf{E}[y^2]|}_{=O(h^\beta)} + \underbrace{|\mathsf{E}[X] + \mathsf{E}[y]|}_{\le \text{const}} \cdot \underbrace{|\mathsf{E}[X] - \mathsf{E}[y]|}_{=O(h^\beta)},\end{aligned}$$

i.e. convergence of the variance.

Remark

Strong convergence implies weak convergence with respect to $g(x) = x$.

Because the properties of integration

$$|\ \mathsf{E}[X] - \mathsf{E}[Y]\ | = |\mathsf{E}[X - Y]| \le \mathsf{E}\left[\ |X - Y|\ \right]$$

lead to

$$\mathsf{E}\left[\ |X - Y|\ \right] = O(h^\gamma) \quad \Longrightarrow \quad \mathsf{E}[X] - \mathsf{E}[Y] = O(h^\gamma)\,.$$

Practical advantage of weak convergence:

The increments ΔW needed to calculate y^h can be replaced by other random variables $\widehat{\Delta W}$ with matching first moments. The weak-convergence order survives.

Example

$\widehat{\Delta W} := \pm\sqrt{\Delta t}$, where both signs occur with probability 1/2. (cheaper to approximate than $Z \sim \mathcal{N}(0,1)$)

This implies $\mathsf{E}(\widehat{\Delta W}) = 0$ and $\mathsf{E}((\widehat{\Delta W})^2) = \Delta t$ $(\Rightarrow \mathrm{Var}(\widehat{\Delta W}) = \Delta t)$.

When $\widehat{\Delta W}$ replaces ΔW one obtains the "simplified Euler method," which is weakly convergent with order 1.

3.2 Constructing Integrators for SDEs

The derivation of integrators for SDEs can be based on the stochastic Taylor expansion.

A. Stochastic Taylor expansion

(follows [P. Kloeden & E. Platen: Numerical Solution of SDEs])
For motivation we first consider the deterministic autonomous case

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = a(X_t)\,.$$

The chain rule for $f \in C^1(\mathbb{R})$ tells

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f(X_t) &= \frac{\mathrm{d}f(X)}{\mathrm{d}X}\cdot\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{\mathrm{d}f(X)}{\mathrm{d}X}a(X_t)\\ &= a(X_t)\frac{\mathrm{d}}{\mathrm{d}X}f(X_t) =: Lf(X_t)\,.\\ \Longrightarrow f(X_t) &= f(X_{t_0}) + \int_{t_0}^{t} \underbrace{Lf(X_s)}_{=:\tilde{f}}\,\mathrm{d}s\end{aligned}$$

Substitute this formula for

$$\tilde{f}(X_s) := Lf(X_s)$$

into itself gives

$$\begin{aligned}f(X_t) =& f(X_{t_0}) + \int_{t_0}^{t}\left\{\tilde{f}(X_{t_0}) + \int_{t_0}^{s} L\tilde{f}(X_z)\,\mathrm{d}z\right\}\mathrm{d}s\\ =& f(X_{t_0}) + \tilde{f}(X_{t_0})\int_{t_0}^{t}\mathrm{d}s + \int_{t_0}^{t}\int_{t_0}^{s} L\tilde{f}(X_z)\,\mathrm{d}z\,\mathrm{d}s\\ =& f(X_{t_0}) + Lf(X_{t_0})(t-t_0) + \int_{t_0}^{t}\int_{t_0}^{s} L^2 f(X_z)\,\mathrm{d}z\,\mathrm{d}s\,.\end{aligned}$$

This is the Taylor expansion with remainder term in integral form, here expanded until the linear term; the remainder is a double integral. This process can be continued, and the deterministic Taylor expansion with remainder in integral form results. (All needed derivatives may exist.)

Now we turn to the **stochastic case**, to the Itô–Taylor expansion:

Applying the Itô lemma on $f(X)$ and the autonomous SDE

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

leads to

$$\mathrm{d}f(X_t) = \underbrace{\{a(X_t)\frac{\partial}{\partial x}f(X_t) + \frac{1}{2}\left(b(X_t)\right)^2\frac{\partial^2}{\partial x^2}f(X_t)\}}_{=:L^0 f(X_t)}\,\mathrm{d}t + \underbrace{b(X_t)\frac{\partial}{\partial x}f(X_t)}_{=:L^1 f(X_t)}\,\mathrm{d}W_t\,,$$

or

$$\boxed{f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} L^0 f(X_s)\,\mathrm{d}s + \int_{t_0}^{t} L^1 f(X_s)\,\mathrm{d}W_s\,.} \qquad (*)$$

Specifically for $f(x) = x$, the equation $(*)$ includes the starting SDE

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

Now apply $(*)$ for suitable $\tilde{f}$, begin with $\tilde{f} := a$ and $\tilde{f} := b$, and obtain

$$X_t = X_{t_0} + \int_{t_0}^{t} \left\{ a(X_{t_0}) + \int_{t_0}^{s} L^0 a(X_z)\,\mathrm{d}z + \int_{t_0}^{s} L^1 a(X_z)\,\mathrm{d}W_z \right\} \mathrm{d}s$$
$$+ \int_{t_0}^{t} \left\{ b(X_{t_0}) + \int_{t_0}^{s} L^0 b(X_z)\,\mathrm{d}z + \int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z \right\} \mathrm{d}W_s\,.$$

This can be written

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

with remainder

$$R = \int_{t_0}^{t} \int_{t_0}^{s} L^0 a(X_z)\,\mathrm{d}z\,\mathrm{d}s + \int_{t_0}^{t} \int_{t_0}^{s} L^1 a(X_z)\,\mathrm{d}W_z\,\mathrm{d}s$$
$$+ \int_{t_0}^{t} \int_{t_0}^{s} L^0 b(X_z)\,\mathrm{d}z\,\mathrm{d}W_s + \int_{t_0}^{t} \int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z\,\mathrm{d}W_s\,.$$

The integrands are

$$L^0 a = aa' + \frac{1}{2}b^2 a'' \qquad L^0 b = ab' + \frac{1}{2}b^2 b''$$
$$L^1 a = ba' \qquad L^1 b = bb'\,.$$

Analogously the integrands in the double integrals in R can be replaced, by applying $(*)$ with proper $\tilde{f}$. Thereby, the double integrals

$$\underbrace{\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}z\,\mathrm{d}s}_{=:\mathcal{I}(0,0)=\frac{1}{2}(\Delta t)^2}\,, \quad \underbrace{\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}s}_{=:\mathcal{I}(1,0)}, \quad \underbrace{\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}z\,\mathrm{d}W_s}_{=:\mathcal{I}(0,1)}\,, \quad \underbrace{\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}W_s}_{=:\mathcal{I}(1,1)}$$

occur as factors. $\mathcal{I}(1,0), \mathcal{I}(0,1), \mathcal{I}(1,1)$ are stochastic variables. By a plausibility argument (replace $\Delta W_s := W_s - W_{t_0}$ by its expectation $\sqrt{s - t_0}$) expect that $I(1,1)$ is the integral of lowest order: $O(\Delta t)$. We begin with this integral, for $\tilde{f} := L^1 b(X)$. From $(*)$ conclude

$$\int_{t_0}^{t} \int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z\,\mathrm{d}W_s = L^1 b(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}W_s + \text{two triple integrals}$$

Then R consists of

$$R = \text{ three double integrals } + \underbrace{b(X_{t_0})b'(X_{t_0})}_{=L^1 b(X_{t_0})}\,\mathcal{I}(1,1) \quad + \text{ two triple integrals}$$

The next issue is to calculate the double integral $\mathcal{I}(1,1)$:

Let $g(x) := x^2$ and $X_t = W_t$, which solves an SDE with $a = 0$ and $b = 1$. The Itô lemma implies

$$\mathrm{d}(W_t^2) = \frac{1}{2} 2 \, \mathrm{d}t + 2W_t \, \mathrm{d}W_t = \mathrm{d}t + 2W_t \, \mathrm{d}W_t,$$

which in turn yields

$$\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}W_z \, \mathrm{d}W_s = \int_{t_0}^{t} (W_s - W_{t_0}) \, \mathrm{d}W_s = \int_{t_0}^{t} W_s \, \mathrm{d}W_s - W_{t_0} \int_{t_0}^{t} \mathrm{d}W_s =$$

$$\int_{t_0}^{t} \tfrac{1}{2} \left[\mathrm{d}(W_s^2) - \mathrm{d}s \right] - W_{t_0}(W_t - W_{t_0}) =$$

$$\tfrac{1}{2}(W_t^2 - W_{t_0}^2) - \tfrac{1}{2}(t - t_0) - \tfrac{2}{2} W_{t_0}(W_t - W_{t_0}) = \tfrac{1}{2}(\Delta W_t)^2 - \tfrac{1}{2}\Delta t \, .$$

This confirms the anticipated order $O(\Delta t)$ of $\mathcal{I}(1,1)$.

The above derivation of the stochastic Taylor expansion can be continued. This calls for a systematic definition and notation of the multi-integrals, for example, $\mathcal{I}(0,0,0), \ldots$. In this notation, a "0" stands for a deterministic integration, and a "1" for a stochastic integration.

Application

Attach further leading terms of the stochastic Taylor expansion to obtain integrators of higher order.

Example (Milstein method)

$$y_{j+1} = y_j + a \, \Delta t + b \, \Delta W_j + \tfrac{1}{2} b b' \left\{ (\Delta W_j)^2 - \Delta t \right\}$$

The first terms represent the Euler method, and the last term completes the list of $O(\Delta t)$-terms, and improves the low order of strong convergence to 1. The weak order is also 1. (This may be checked empirically.) Question: What does this result mean in view of SDEs with $b' = 0$?

B. Positivity

As mentioned before, positive solutions are characteristic for many SDEs in finance. This should be preserved by numerical approximations. We discuss this topic for the CIR process, which is part of the Heston model.

Example CIR

$$\mathrm{d}X_t = \kappa(\theta - X_t) \, \mathrm{d}t + \sigma \sqrt{X_t} \, \mathrm{d}W_t$$

with $\kappa, \theta, \sigma > 0$, $X_0 = x_0 > 0$. Positivity of X_t for all t is established by the "Feller condition"

$$\kappa\theta \geq \frac{1}{2}\sigma^2 \, ,$$

which guarantees a strong enough growth rate. We remark in passing that $b(X) = \sigma\sqrt{X}$ does not satisfy a global Lipschitz condition.

Euler scheme:

$$y_{j+1} = y_j + \kappa(\theta - y_j)\Delta t + \sigma \sqrt{y_j} \, \Delta W_j$$

with $y_0 := x_0$, works as long as $y_j \geq 0$. There is a positive probability that y_{j+1} is negative. When X represents an asset price, an interest rate, or a variance (Heston model), then $y < 0$ must be avoided.

Variants

For example, replace $\sqrt{y}$ by $\sqrt{|y|}$ or by $\sqrt{y^+}$. Then the scheme is defined for all $y \in \mathbb{R}$. Another variant calculates

$$y_{j+1} = | \, y_j + \kappa(\theta - y_j)\Delta t + \sigma\sqrt{y_j}\,\Delta W_j \, | \, .$$

Implicit Euler methods can be applied as well, for example, the drift-implicit scheme

$$y_{j+1} = y_j + a(y_{j+1})\Delta t + b(y_j)\,\Delta W_j \, .$$

If this scheme is applied to the SDE of the square root process $\sqrt{X_t}$, then a quadratic equation for y_{j+1} results with a unique positive solution (Exercise !). [A. Alfonsi (2005)]

3.3 Monte Carlo Methods for European Options

The aim is to calculate the value

$$V(S_0, 0) = \mathrm{e}^{-rT}\mathsf{E}_\mathsf{Q}\,[\,\Psi(S_T) \mid S(0) = S_0\,]$$

of a European option, where Ψ denotes the payoff and Q a risk-free probability measure.

A. Basic Principle

The integral of this expectation can be approximated by Monte Carlo methods. The first decision is the choice of the market model (as Heston- or Black–Scholes model). Here we focus on the classic Black-Scholes model with GBM,

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

The procedure is analogous to

Monte Carlo quadrature:

With respect to $\mathcal{U}[0,1]$,

$$\int_0^1 f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathbf{1}_{[0,1]}\,\mathrm{d}x = \mathsf{E}(f)\,.$$

Applying the law of large numbers,

$$\frac{1}{N}\sum_{k=1}^{N} f(x_k) \longrightarrow \mathsf{E}(f) \quad \text{for } N \to \infty\,,$$

where x_k are independent random uniformly distributed numbers in the domain $\mathcal{D} := [0,1]$, because $\mathbf{1}_{[0,1]}$ is the corresponding density. Hence the sum

$$\frac{1}{N}\sum_{k=1}^{N} f(x_k)$$

approximates the integral $\int_0^1 f(x)\,\mathrm{d}x$.

For general domains $\mathcal{D}$ the approximation is

$$\int_{\mathcal{D}} f(x)\,\mathrm{d}x \approx \frac{\mathrm{Vol}(\mathcal{D})}{N} \sum_{k=1}^{N} f(x_k)\,.$$

The **error** is probabilistic. The central limit theorem provides related assertions, for example:

With 95% probabilty the true value of the integral lies in the confidence interval around the approximate value, which is given by the half width $a\sigma/\sqrt{N}$. For 95% probability the parameter a is $a = 1.96$, and σ is the standard deviation.

So much on MC applied to quadrature. Now the question is, what is the structure of f when options are to be priced under GBM?

In that case, the density is f_{GBM}, and accordingly the x_k must be distributed lognormally.

Algorithm: Monte Carlo Method for European Options

Simulate N paths of the asset price under the risk-neutral measure Q. Each path starts at S_0, and terminates in $x_k := (S_T)_k$ for $k = 1, ..., N$. Evaluate the payoff $\Psi(S_T)$

$$f(x_k) := \Psi\big((S_T)_k\big)\,,$$

calculate the mean, and discount with factor e^{-rT}. Its expectation yields the true value V as long as f (S_T and Ψ) is unbiased.

(Illustration in the beginning of the chapter)

Examples of a payoff Ψ

1.) *Binary* or *digital* option, e.g., binary call:

$$\Psi(S_T) = \mathbf{1}_{S_T > K} = \begin{cases} 1 & \text{in case } S_T > K \\ 0 & \text{elsewhere} \end{cases}$$

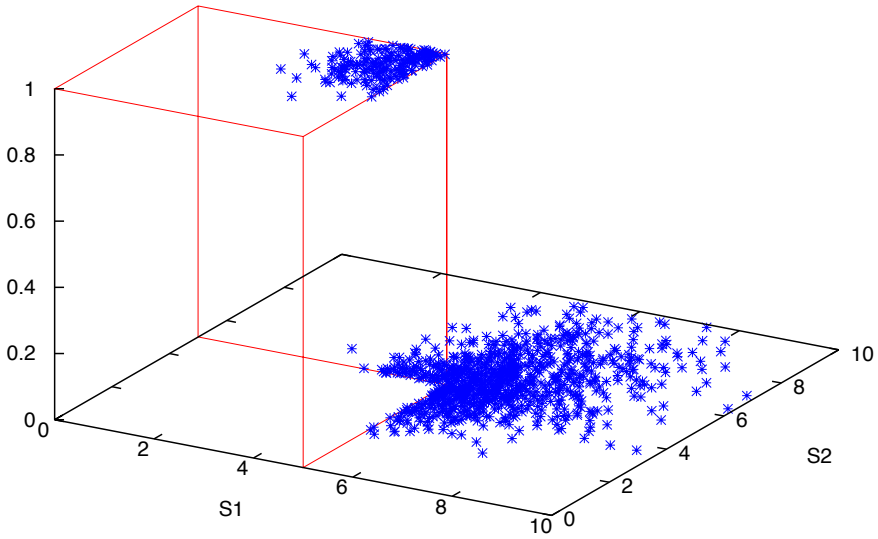
2.) *Barrier* option with barrier B, e.g. a *down-and-out* call option with

$$\Psi(S) = \begin{cases} 0 & \text{in case } S_t \le B \text{ for a } t \text{ in } 0 \le t \le T \\ (S_T - K)^+ & \text{elsewhere} \end{cases}$$

For this path-dependent exotic option the entire path S_t on $0 \le t \le T$ is of interest. (meaningful for $S_0 > B > K$; illustration in (S, t, V)-space under consideration of boundary conditions along $S = B$)

3.) *two-asset cash-or-nothing* put: The payoff is 1 in case the inequalities $S_1(T) < K_1$ and $S_2(T) < K$ hold, where $S_1(t), S_2(t)$ denote the prices of the two assets. (See the figure below; more figures in www.compfin.de.)

Hint: Many analytic solution formulas can be found in [E.G. Haug: Option Pricing Formulas].



Implementation of the Monte Carlo Method

For the GBM model the true solution

$$S_t = S_0 \exp\left\{(r - \tfrac{1}{2}\sigma^2)t + \sigma W_t\right\}$$

can be applied. For options that are not path-dependent this requires only one random number for each path, for generating W_T and S_T. For more general models without analytic solution formula, one must resort to numerical integration (say, with Euler's method). Then Monte Carlo consists of two loops: the outer loop of sampling ($k = 1, \ldots, N$), and the inner loop of the integration ($j = 1, \ldots, M$; $\Delta t = \frac{T}{M}$; $t_j = j\Delta t$). For path-dependent GBM models, the analytic formula can be applied in a piecewise fashion,

$$S_{t_{j+1}} = S_{t_j} \exp\left\{(r - \tfrac{1}{2}\sigma^2)\,\Delta t + \sigma\,\Delta W\right\}$$

for all j, with $\Delta W = \sqrt{\Delta t}\,Z$, $Z \sim \mathcal{N}(0,1)$.

Dimension. Monte Carlo works in the same way for high-dimensional problems. The costs are essentially independent on the dimension. This is an important advantage of Monte Carlo methods.

B. Accuracy

a) Denote

$$\hat{\mu} := \frac{1}{N}\sum_{k=1}^{N} f(x_k)\,, \quad \hat{s}^2 := \frac{1}{N-1}\sum_{k=1}^{N}\left(f(x_k) - \hat{\mu}\right)^2\,,$$

and $\mu = \mathsf{E}(\hat{\mu})$. According to the **central limit theorem**, the approximation $\hat{\mu}$ obeys $\mathcal{N}(\mu, \sigma^2)$,

$$\mathsf{P}\left(\hat{\mu} - \mu \le a\frac{\sigma}{\sqrt{N}}\right) = F(a),$$

with distribution function F. In practice σ^2 is replaced by its approximation $\hat{s}^2$. The error behaves as $\frac{\hat{s}}{\sqrt{N}}$. To reduce this statistical error, either reduce the numerator *(variance reduction)*, or enlarge the denominator. The latter means to increase the number of simulations,

and is very costly. For example, to gain one additional correct decimal, the error must be reduced by a factor $\frac{1}{10}$, which amounts to raise the costs by a factor of $100 = (\frac{1}{10})^{-2}$.

b) In several cases, the computation of $f(x_i)$ gives rise to another error, namely, the **bias**. Let $\hat{x}$ be an estimator of the true value x to be estimated, then the bias is defined as

$$\text{bias}(\hat{x}) := \mathsf{E}[\hat{x}] - x.$$

Examples

1.) For a *lookback* option the payoff involves the variable

$$x := \mathsf{E}\left[\max_{0\le t\le T} S_t\right] .$$

An approximation is

$$\hat{x} := \max_{0\le j\le m} S_{t_j} .$$

Clearly $\hat{x} \le x$. Almost surely $\hat{x}$ underestimates x, i.e. $\mathsf{E}[\hat{x}] < x$. Hence $\text{bias}(\hat{x}) \neq 0$.

2.) Compared to the analytic solution of GBM, the Euler method provides biased results. For GBM,

$$S_{t_{j+1}} = S_{t_j} \exp\left\{(r - \tfrac{1}{2}\sigma^2)\,\Delta t + \sigma\,\Delta W\right\}$$

is unbiased, whereas the Euler step

$$S_{t_{j+1}} = S_{t_j}(1 + r\,\Delta t + \sigma\,\Delta W)$$

is biased.

These two examples are asymptotically unbiased since the bias vanishes for $M \to \infty$. To reduce the errors, there are several possibilities, which must be compared for costs and tradeoffs. Either

- apply variance reduction,
- increase N, or
- reduce the bias (M larger, Δt smaller),

or apply all these measures. Increasing N and M should be balanced. The overall error is measured by the mean square error:

Definition (mean square error)

$$\text{MSE}(\hat{x}) := \mathsf{E}\left[(x - \hat{x})^2\right] .$$

As is easily verified,

$$\begin{aligned}\text{MSE}(\hat{x}) &= (\mathsf{E}[\hat{x}] - x)^2 + \mathsf{E}\left[(\hat{x} - \mathsf{E}(\hat{x}))^2\right] \\ &= (\text{bias}(\hat{x}))^2 + \mathsf{Var}(\hat{x}) .\end{aligned}$$

C. Variance Reduction

There are several methods of variance reduction. The simplest (and maybe the least powerful) is the method of *antithetic variates*. As for the crude MC, paths are simulated with random numbers $Z_1, Z_2, \ldots$ Let us denote the MC approximation $\hat{V}$. The idea of antithetic variates is to use in parallel the numbers $-Z_1, -Z_2, \ldots$, which are also $\sim \mathcal{N}(0,1)$, to calculate "mirror paths" S_t^- from which the payoff values $\Psi(S_T^-)$ are calculated. This leads to a second Monte Carlo value V^-. By construction, $\mathsf{Var}(\hat{V}) = \mathsf{Var}(V^-)$. The effort to calculate V^- is slightly lower than that for $\mathsf{Var}(\hat{V})$ because the Z's are recycled. The mean

$$V_{\rm AV} := \frac{1}{2}(\hat{V} + V^-)$$

satisfies

$$\begin{aligned}\mathsf{Var}(V_{\rm AV}) &= \frac{1}{4}\mathsf{Var}(\hat{V} + V^-) \\ &= \frac{1}{4}(\mathsf{Var}\hat{V} + \mathsf{Var}V^- + 2\mathsf{Cov}(\hat{V}, V^-)) \\ &= \frac{1}{2}\mathsf{Var}\hat{V} + \frac{1}{2}\mathsf{Cov}(\hat{V}, V^-)\end{aligned}$$

The anti-symmetric construction of the mirror paths inspires some confidence that the results are negatively correlated, $\mathsf{Cov}(\hat{V}, V^-) < 0$. This holds in case the dependence of the output V on the input Z is monotonic. For $\mathsf{Cov}(\hat{V}, V^-) < 0$ the effect is

$$\mathsf{Var}(V_{\rm AV}) < \frac{1}{2}\mathsf{Var}(\hat{V})\,.$$

This approach at most doubles the costs. In comparison, an error reduction of this size (factor $< \frac{1}{2}$) by merely increasing N requires at least fourfold costs.

Example GBM: Let the index k in V_k label the MC simulation, $k = 1, \ldots, N$, for GBM. For a payoff Ψ draw $Z_k \sim \mathcal{N}(0,1)$ and calulate the pairs $\hat{V}_k, V_k^-$ and the antithetic variate $V_{{\rm AV},k}$ as follows (for $t_0 = 0$):

$$\begin{aligned}\hat{V}_k &= \Psi\left(S_0 \exp\left\{(r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}\,Z_k\right\}\right) \\ V_k^- &= \Psi\left(S_0 \exp\left\{(r - \frac{\sigma^2}{2})T - \sigma\sqrt{T}\,Z_k\right\}\right) \\ V_{{\rm AV},k} &= \frac{1}{2}(\hat{V}_k + V_k^-)\end{aligned}$$

For each k, $\hat{V}_k$ and V_k^- are dependent, but the independence of $Z_k \sim \mathcal{N}(0,1)$ makes the $V_{{\rm AV},k}$ for $k = 1, \ldots, N$ independent, and MC is applied: The mean, discountend with factor e^{-rT}, approximates V.

(For the more complex method of *control variates*, see the literature.)

Notice that MC has not been developed for the simple vanilla options. The potential of MC is needed for exotic options, in particular, in high-dimensional situations.

3.4 Monte Carlo Methods for American Options

A. Stopping Time

Examples of decisions in the financial market include selling an asset, or exercising an American option. Let us call the decision "stopping," and the time instant of the decision "stopping time" τ. Such decisions can only be based on the information available so far. Accordingly, a stopping time must be **non-anticipating**: That is, for any time t one must know whether the decision is made, i.e. whether $\tau \le t$ or $\tau > t$.

This characterization of a stopping time can be defined formally with the means of stochastics, building on the underlying process S_t:

> Recall the filtration $\mathcal{F}_t$: A stochastic process S_t is called $\mathcal{F}_t$-adapted, if S_t is $\mathcal{F}_t$-measurable for all t. The natural filtration $\mathcal{F}_t^S$ is the smallest sigma-algebra over $\{S_s \mid 0 \le s \le t\}$, augmented by the P-null sets. S_t is $\mathcal{F}_t^S$-adapted. Filtrations represent the amount of information available at time t. Hence we require for the set $\{\tau \le t\}$ of all decisions until t
> $$\{\tau \le t\} \in \mathcal{F}_t\,,$$
> which is the $\mathcal{F}_t$-measurability of τ.

Definition (stopping time)

> A stopping time τ with respect to a filtration $\mathcal{F}_t$ is a random variable with values in $[0,T] \cup \{\infty\}$, which is $\mathcal{F}_t$-measurable for all t.

The importance of stopping times for American-style options is highlighted by the following result of Bensoussan (1984):

Let $\Psi(S_t)$ be a payoff, e.g. $\Psi(S_t) = (K - S_t)^+$.
The value of an American option is

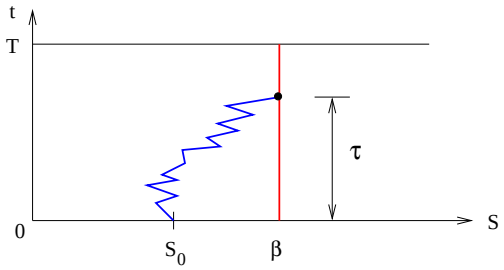
$$V(S,0) = \sup_{\substack{0 \le \tau \le T \\ \tau \text{ stopping time}}} \mathsf{E}_{\mathsf{Q}}\left[\, \mathrm{e}^{-r\tau}\, \Psi(S_\tau) \mid S_0 = S \,\right] \qquad (*)$$

The stopping time τ is with respect to a natural filtration $\mathcal{F}_t$ of S_t.

Examples of stopping times

1) Define the *hitting time*
$$\tau := \inf\{\, t > 0 \mid S_t \ge \beta \,\}$$
for given $\beta > S_0$. If no such t exists, set $\tau := \infty$.
Clearly, a hitting time is non-anticipating: Setting τ amounts to setting a flag in case of hitting, and for any time t check whether the flag is set.*

* See [Hunt & Kennedy (2000)] for a formal proof that this τ is stopping time.



2) Define t^* as the time instant at which $\max_{0\le t\le T} S_t$ is reached. This is no stopping time! Because for arbitrary t one can not decide whether $t^* \le t$ or $t^* > t$.

3) A stopping time is given by

$$\tau := \min\{\, t \le T \mid (t, S_t) \in \textit{stopping region}\,\}$$

(This is a hitting time for S_t hitting the early-exercise curve; cf. Section 4.5.)

B. Parametric Methods

In $(*)$ the supremum over *all* stopping times is taken. Now we construct a special stopping time. Similar as in Example 1 or 3 we define a curve in the (S,t)-half strip, which is supposed to approximate the early-exercise curve. This defines a special stopping strategy $\tilde\tau$ by the event of hitting the curve. Assume that β is a vector of parameters defining the curve. Then the stopping rule and $\tilde\tau$ depend on β. This special β-depending stopping strategy $\tilde\tau$ leads to a **lower bound**

$$V^{\mathrm{low}(\beta)}(S,0) := \mathsf{E_Q}\left[\, \mathrm{e}^{-r\tilde\tau}\Psi(S_{\tilde\tau}) \mid S_0 = S \,\right] \le V(S,0)\,.$$

Application: Obviously, $V(S,0)$ can be approximated via suitable β-defined stopping curves as

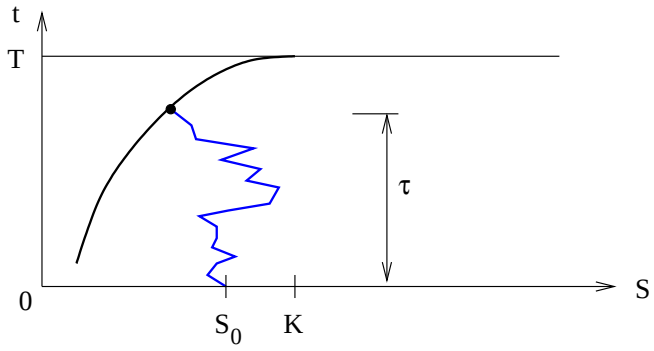
$$\sup_\beta V^{\mathrm{low}(\beta)}.$$

This idea of an optimal stopping strategy leads to the procedure:

> Construct a curve depending on a parameter vector β such that the curve approximates the early-exercise curve. The stopping strategy is to stop when the path S_t crosses the curve defined by β. For N such paths evaluate the payoff, and evaluate (approximate) the value $V^{\mathrm{low}(\beta)}$ as crude MC does. Next attempt to maximize the lower bound $V^{\mathrm{low}(\beta)}$ by repeating the procedure for "better" parameters β.

Example with $\beta \in \mathbb{R}^1$:

> Consider a parabola with peak in $(S,t) = (K,T)$, which is defined by one parameter β only. As illustrated in the figure, this parabola can be seen as an approximation of the early-exercise curve of a put. Calculate N paths (e.g. $N = 10000$) until the left branch of the parabola (yields $\tilde\tau$) or $t = T$ is reached. Similar as in Example 1 this hitting time gives rise to an approximation $V^{\mathrm{low}(\beta)}$. Each evaluation of $V^{\mathrm{low}(\beta)}$ costs as much as MC for a European option. Then repeat the procedure with a better β. — This $V^{\mathrm{low}(\beta)}$ will not converge to $V(S,0)$. A systematic error will remain because the eary-exercise curve can not be approximated so well with a simple parabola.



To complete the procedure, one should also construct an upper bound $V^{\rm up}$. Literature: [P. Glasserman: Monte Carlo Methods in Financial Engineering (2004)]

C. Regression Methods

Definition (Bermudan option)

A Bermudan option is an option that can be exercised only at a finite number M of discrete time instances t_j.

Specifically for

$$t_j := j\,\Delta t\,, \quad \Delta t := \frac{T}{M} \quad (j = 0, \ldots, M)$$

we denote the value of a Bermudan option $V^{\rm Be(M)}$. Because of the additional exercise possibilities,

$$V^{\rm Eu} \le V^{\rm Be(M)} \le V^{\rm Am}$$

holds. One can show

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

For suitable M the value $V^{\rm Be(M)}$ is used as approximation of $V^{\rm Am}$. The linear convergence suggests working with a few moderate values of M and apply Richardson extrapolation. In this way, the high costs of Monte Carlo for American options can be kept at a tolerable level.

Recall from the binomial method (Section 1.3): The value of an American option is calculated recursively in a backward fashion, where the continuation values $V^{\rm cont}$ are defined as European options on a strip, and for each t_j

$$V^{\rm Am} = \max\left\{\Psi(S), V^{\rm cont}\right\}\,.$$

Because at each t_j the holder of the option decides which of the two possibilities {exercise, hold} is optimal.*

For a Bermudan option we define the continuation value at t_j analogously:

$$C_j(x) := {\rm e}^{-r\Delta t}\,\mathsf{E}_{\mathsf{Q}}\left[V(S_{t_{j+1}}, t_{j+1}) \mid S_{t_j} = x\right].$$

These functions $C_j(x)$ must be approximated.

---

* principle of dynamic programming

General Recursion

Set $V_M(x) \equiv \Psi(x)$.
For $j = M-1, \ldots, 1$
construct $C_j(x)$ for $x > 0$;
$V_j(x) := V(x, t_j) = \max\{\Psi(x),\, C_j(x)\}$ for grid points x.
$V_0 := V(S_{t_0}, t_0) = \max\{\, \Psi(S_0),\, C_0(S_0)\,\}$.

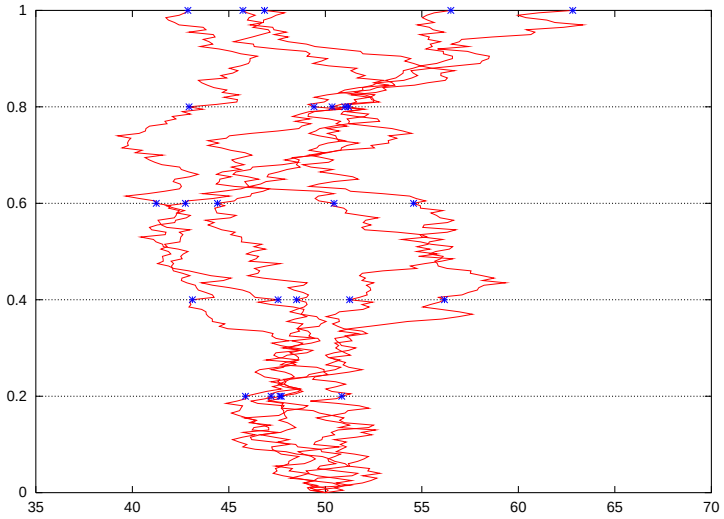
Below we define special x by a stochastic grid.

To calculate the functions $C_j(x)$ with Monte Carlo, one draws information out of paths established by simulation, and approximates $C_j(x)$ by a regression curve $\hat{C}_j(x)$.

Regression (basic version)

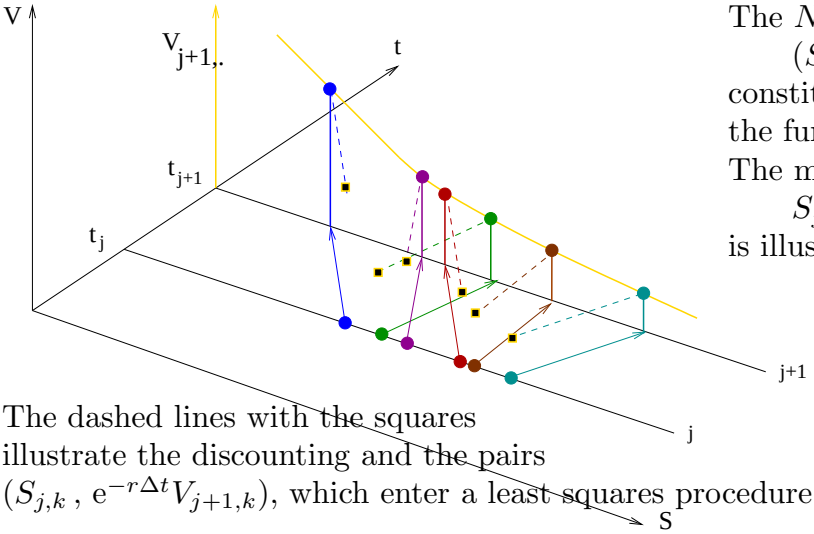
(a) Simulate N paths $S_1(t), ..., S_N(t)$: Calculate and store the values

$$S_{j,k} := S_k(t_j)\,, \quad j = 1, ..., M,\ k = 1, ..., N\,.$$



five trajectories and points $(S_{j,k}, t_j)$ for $j = 1, \ldots, 5$, $k = 1, \ldots, 5$

Mapping C, illustration for $N = 6$



The N pairs $(k = 1, \ldots, N)$
$(S_{j,k}\,,\ \mathrm{e}^{-r\Delta t} V_{j+1,k})$
constitute the data needed to calculate the function $\hat{C}_j(x)$.
The mapping
$S_{j,k} \longrightarrow V_{j+1,k}$
is illustrated by circular dots.

The dashed lines with the squares illustrate the discounting and the pairs $(S_{j,k}\,,\ \mathrm{e}^{-r\Delta t} V_{j+1,k})$, which enter a least squares procedure to generate $\hat{C}_j(x)$.

Regression (continue)

(b) For $j = M$ set $V_{M,k} = \Psi(S_{M,k})$ for all k.

(c) For $j = M-1, ..., 1$:
Approximate $C_j(x)$ with suitable base functions $\phi_0, ..., \phi_L$ (e.g. monomials)

$$C_j(x) \approx \sum_{l=0}^{L} a_l \phi_l(x) =: \hat{C}_j(x)$$

To this end, apply least-squares minimization on the N points

$$(S_{j,k}\,,\, \mathrm{e}^{-r\Delta t} V_{j+1,k})\ , \quad k = 1, ..., N$$

to get the coefficients $a_0, \ldots, a_L$ and thus $\hat{C}_j$.
Evaluation:

$$V_{j,k} := \max\left\{ \Psi(S_{j,k}),\, \hat{C}_j(S_{j,k}) \right\} .$$

(d) Set

$$V_0 := \max\{\, \Psi(S_0),\, \mathrm{e}^{-r\Delta t} \frac{1}{N} (V_{1,1} + ... + V_{1,N}) \,\} .$$

Costs

The expensive steps are (a) and (c). Step (d) is needed because the algorithm of (c) can not be applied for $j = 0$, since $S_{0,k} = S_0$ for all k. Instead, the mean in (d) is taken. Convergence of the algorithm was proved.

Based on this regression framework, the algorithm of Longstaff & Schwartz (2001) is built, as well as the even more efficient algorithm by C. Jonen (2009). In particular, step (c) offers potential for improvements.

The costs of step (c) of the above algorithm for American options do depend on the dimension. (Why?) When the entire computing time for pricing an option is limited, then the dependence on the dimension restricts the achievable accuracy. In this sense, the error of MC depends on the dimension.

Illustration: Topic 6

Supplement to Section 3.4

Longstaff & Schwartz modify the algorithm as follows:

> A dynamical-programming principle is incorporated for the optimal stopping times. Each path has its own stopping time τ_k for $k = 1, ..., N$. (It suffices to store the index k since $\tau_k = k\Delta t$.) This algorithm takes advantage of the possibility to work across several time levels. Due to a modification of C. Jonen [Intern. J. Computer Math. **86** (2009); PhD 2011] this is an efficient method.
>
> **Algorithm**
>
> Initialization: $\tau_k := M$ for all k.
>
> For each $j = M - 1, ..., 1$:
>
> loop over all paths $k = 1, ..., N$:
>
> In case $\Psi(S_{j,k}) \geq \hat{C}_j(S_{j,k})$ set $\tau_k := j$.
>
> Otherwise leave τ_k unchanged.

For further hints on regression, and on the computation of sensitivities (Greeks), consult [R.U. Seydel: Tools for Computational Finance, Springer, London (2017)]. See also Topic 6 in the *Topics for CF*.

4. Finite Difference Methods for American Vanilla Options

This chapter considers options with vanilla payoff $\Psi(S)$, where S is the price of an underlying asset. The assumed model for S_t is GBM. A continuous flow of dividend payments is admitted, with constant *dividend rate* $\delta \geq 0$. Hence the GBM is

$$\frac{\mathrm{d}S_t}{S_t} = (\mu - \delta)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\,.$$

4.1 Preparations

The Black–Scholes equation for the value $V(S,t)$ of a standard option is

$$\frac{\partial V}{\partial t} + \frac{1}{2}\,\sigma^2\,S^2\,\frac{\partial^2 V}{\partial S^2} + (r-\delta)\,S\,\frac{\partial V}{\partial S} - r\,V = 0\,,$$

defined on the domain $S > 0$, $0 \leq t \leq T$. Terminal condition: $V(S,T) = \Psi(S)$.

This partial differential equation (PDE) can be solved directly with numerical methods. But then tricky stability issues must be tackled. Here we prefer applying transformations as much as possible in order to obtain simpler equations, e.g. with the transformation $S = \mathrm{e}^x$.

Assumption: σ, r and δ are constant.

Then the Black–Scholes equation can be transformed to a strikingly simple type of PDE (exercises). With $t = T - \frac{2\tau}{\sigma^2}$, $S = K\mathrm{e}^x$, $q := \frac{2r}{\sigma^2}$, $q_\delta := \frac{2(r-\delta)}{\sigma^2}$ and $y(x,\tau)$ defined by

$$V(S,t) = V(K\mathrm{e}^x, T - \frac{2\tau}{\sigma^2}) =: v(x,\tau)$$
$$v(x,\tau) = K\exp\left\{-\frac{1}{2}(q_\delta - 1)x - \left(\frac{1}{4}(q_\delta-1)^2 + q\right)\tau\right\} y(x,\tau)$$

one obtains

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

This is a parabolic equation which has an analytic solution. A back transformation establishes the well-known Black–Scholes *formula*. So, for a standard option of European style no specific numerical method is needed.

The aim of the chapter is to price **American** options.

Generic applications of the simple PDE $y_\tau = y_{xx}$ are heat conduction and diffusion in a one-dimensional medium. The problem is *well-posed* as **initial-value problem** with increasing τ. The time transformation $t \to \tau$ converts the terminal condition for $V(S,T) = \Psi(S)$ to an initial condition for $\tau = 0$:

$$\begin{aligned} \text{call}: \quad & y(x,0) = \max\left\{ \mathrm{e}^{\frac{x}{2}(q_\delta+1)} - \mathrm{e}^{\frac{x}{2}(q_\delta-1)}, 0 \right\} \\ \text{put}: \quad & y(x,0) = \max\left\{ \mathrm{e}^{\frac{x}{2}(q_\delta-1)} - \mathrm{e}^{\frac{x}{2}(q_\delta+1)}, 0 \right\} \end{aligned}$$

A solution y of the initial-value problem is defined on the domain

$$-\infty < x < +\infty, \quad 0 \le \tau \le \frac{1}{2}\sigma^2 T =: \tau_{\max}\,,$$

which is a strip in the (x,τ)-plane. "Boundary" here means $x \to -\infty$, $x \to +\infty$.

Boundary-Value Problem

For numerical purposes the infinite strip is truncated to a rectangle, with

$$x_{\min} \le x \le x_{\max}\,.$$

This "localization" cuts down the influence of the initial conditions, and additional conditions must be formulated to make the problem well-posed. These are boundary conditions for the sides $x_{\min}$ and $x_{\max}$. For the solution $w(x,\tau)$ of the **boundary-value problem** on the rectangle we aim at $w \approx y$, which requires the rectangle to be "large." The sides $x_{\min}$ and $x_{\max}$ must be chosen accordingly. In addition, the interval must include the range of financial interest, namely, the x-values of S_0 and K. This requires

$$x_{\min} < \min\left\{0,\, \log\frac{S_0}{K}\right\}\,, \quad \max\left\{0,\, \log\frac{S_0}{K}\right\} < x_{\max}\,.$$

For simplicity, just think of $x_{\min} = -3$ and $x_{\max} = 3$. — Later we shall see that for American options the partial differential equation mutates to an inequality.

4.2 Basics of Finite-Difference Methods (FDM)

A. Difference Approximations

Recall

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{h}{2} f''(\xi) \quad \text{for a } \xi \in (x, x+h) \text{ and } f \in C^2$$

We introduce an equidistant grid with grid points x_i

$$\ldots < x_{i-1} < x_i < x_{i+1} < \ldots$$

and $h := x_{i+1} - x_i$. [In the non-equidistant case one often prefers finite-element methods (FEM)]. This chapter is confined to equidistant grids.

Analogously, with notation $f_i = f(x_i)$, the following holds true:

$$\begin{aligned} f'(x_i) &= \frac{f_{i+1} - f_{i-1}}{2h} + O(h^2) && \text{for } f \in C^3 \\ f''(x_i) &= \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + O(h^2) && \text{for } f \in C^4 \\ f'(x_i) &= \frac{-f_{i+2} + 4f_{i+1} - 3f_i}{2h} + O(h^2) && \text{for } f \in C^3 \end{aligned}$$

B. The Grid

For an $m \in \mathbb{N}$ and a $\nu_{\max} \in \mathbb{N}$ define

$$\begin{aligned} \Delta x &:= \frac{x_{\max} - x_{\min}}{m}\,, \qquad x_i := x_{\min} + i \cdot \Delta x\,, \qquad i = 0, 1, ..., m\,; \\ \Delta \tau &:= \frac{\tau_{\max}}{\nu_{\max}}\,, \qquad \tau_\nu := \nu \cdot \Delta \tau\,, \qquad \nu = 0, ..., \nu_{\max}\,. \end{aligned}$$

$y_{i,\nu} := y(x_i, \tau_\nu)$ is the value of y at the node (x_i, τ_ν). Approximations of $y_{i,\nu}$ are denoted $w_{i,\nu}$. The values

$$\boxed{w_{i,0} = y(x_i, 0)}$$

are known from the initial conditions.

C. Explicit Method

In the PDE, replace

$$\begin{aligned} \frac{\partial y_{i,\nu}}{\partial \tau} &:= \frac{\partial y(x_i, \tau_\nu)}{\partial \tau} = \frac{y_{i,\nu+1} - y_{i,\nu}}{\Delta \tau} + O(\Delta \tau) \quad \text{and} \\ \frac{\partial^2 y_{i,\nu}}{\partial x^2} &= \frac{y_{i+1,\nu} - 2y_{i,\nu} + y_{i-1,\nu}}{\Delta x^2} + O(\Delta x^2)\,, \end{aligned}$$

drop the O-error terms, replace $y \to w$, and obtain the difference equation

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta \tau} = \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}\,.$$

In case all values w are calculated for the time level ν, then the values of the time level $\nu+1$ are given by

$$w_{i,\nu+1} = w_{i,\nu} + \frac{\Delta \tau}{\Delta x^2}(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu})\,.$$

With the notation

$$\lambda := \frac{\Delta\tau}{\Delta x^2}$$

this is rewritten as

$$\boxed{w_{i,\nu+1} = \lambda w_{i-1,\nu} + (1-2\lambda) w_{i,\nu} + \lambda w_{i+1,\nu}\,.} \qquad (*)$$

Start with $\nu = 0$, since there the $w_{i,0}$ are known. That is, for $\nu = 1$ all $w_{i,1}$ can be calculated with the explicit formula $(*)$, and similarly the $w_{i,\nu}$-values of the following time levels.

ν is the outer loop index and i the inner index. This suggests to describe the procedure by vectors and matrices: Use

$$w^{(\nu)} := (w_{1,\nu}, \ldots, w_{m-1,\nu})^{tr}$$

(so far, reasonable boundary components $w_{0,\nu}$ and $w_{m,\nu}$ are lacking) for the values of the ν-th time level, and the $(m-1)\times(m-1)$-matrix

$$A := A_{\text{expl}} := \begin{pmatrix} 1-2\lambda & \lambda & 0 & \cdots & 0 \\ \lambda & 1-2\lambda & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \lambda \\ 0 & \cdots & 0 & \lambda & 1-2\lambda \end{pmatrix}$$

Now the explicit method can be written

$$\boxed{w^{(\nu+1)} = A w^{(\nu)} \quad \text{for } \nu = 0, 1, 2, ..., \nu_{\max} - 1}$$

up to a modification taking care of boundary conditions for $w_{0,\nu}$ and $w_{m,\nu}$. Preliminarily, for ease of presentation, we have set the boundary conditions to zero.

Example of Instability

$y_\tau = y_{xx}$ with $y(x,0) = \sin \pi x$, $x_0 = 0$, $x_m = 1$ and boundary conditions $= 0$.
Let us approximate $y(x = 0.2, \tau = 0.5)$ using a grid with $\Delta x = 0.1$, i.e. $m = 10$ and $0.2 = x_2$, and two different values of $\Delta\tau$:

a) $\Delta\tau = 0.0005 \quad \Rightarrow \quad \lambda = 0.05 \quad$ and $\quad 0.5 = \tau_{1000}$
yields $w_{2,1000} = 0.00435 \quad$ (exact is 0.004227)

b) $\Delta\tau = 0.01 \quad \Rightarrow \quad \lambda = 1 \quad$ and $\quad 0.5 = \tau_{50}$
yields $w_{2,50} = -1.5 \cdot 10^8 \quad$ (an instability)

D. Stability

Error analysis of $w^{(\nu+1)} = A w^{(\nu)} + d^{(\nu)} \quad$ (above $d^{(\nu)} = 0$)

Notation

$w^{(\nu)}$ exact vectors of $w^{(\nu+1)} = A w^{(\nu)} + d^{(\nu)}$,

$\overline{w}^{(\nu)}$ versions calculated in the computer, subjected to rounding errors,

$e^{(\nu)} := \overline{w}^{(\nu)} - w^{(\nu)}$.

In exact computation, $\overline{w}^{(\nu+1)} - A\overline{w}^{(\nu)} - d^{(\nu)}$ is nonzero; we call this vector $r^{(\nu+1)}$,

$$\overline{w}^{(\nu+1)} = A\overline{w}^{(\nu)} + d^{(\nu)} + r^{(\nu+1)} .$$

The vector $r^{(\nu+1)}$ represents the rounding errors in the ν-th step.

It suffices to study the propagation of *one* error. So we set $r^{(\nu)} = 0$ for $\nu > 1$, i.e., study propagation of the error $e^{(0)}$ in the course of the iterations.

$$\begin{aligned}
&\Rightarrow\ \overline{w}^{(\nu+1)} = A\overline{w}^{(\nu)} + d^{(\nu)} \quad (\nu > 1) \\
&\Rightarrow\ Ae^{(\nu)} = A\overline{w}^{(\nu)} - Aw^{(\nu)} = \overline{w}^{(\nu+1)} - w^{(\nu+1)} = e^{(\nu+1)} \\
&\Rightarrow\ e^{(\nu)} = A^{\nu} e^{(0)}
\end{aligned}$$

For stable behavior, we require $A^{\nu}e^{(0)} \to 0$ for $\nu \to \infty$.

Notation:

$$\rho(A) := \max_k |\mu_k^A| \text{ where } \mu^A \text{ is eigenvalue of } A$$

Lemma 1

$$\rho(A) < 1 \iff A^{\nu} z \to 0 \ \text{ for all } z \text{ and } \nu \to \infty$$

Proof: Textbooks on Numerical Analysis.

A has the structure

$$A = I - \lambda \cdot \underbrace{\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix}}_{=:G} .$$

$\Rightarrow \mu^A = 1 - \lambda\mu^G$.

Lemma 2

$$\text{For } G = \begin{pmatrix} \alpha & \beta & & & 0 \\ \gamma & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta \\ 0 & & & \gamma & \alpha \end{pmatrix} \in \mathbb{R}^{N\times N} \text{ the eigenvalues are}$$

$$\mu_k^G = \alpha + 2\beta\sqrt{\frac{\gamma}{\beta}}\cos\frac{k\pi}{N+1} \ , \ k = 1, ..., N.$$

Proof:

With the eigenvectors

$$v^{(k)} = \left(\sqrt{\frac{\gamma}{\beta}} \sin \frac{k\pi}{N+1}, \left(\sqrt{\frac{\gamma}{\beta}}\right)^2 \sin \frac{2k\pi}{N+1}, ..., \left(\sqrt{\frac{\gamma}{\beta}}\right)^N \sin \frac{Nk\pi}{N+1} \right)^{tr}$$

check $Gv^{(k)} = \mu^G v^{(k)}$.

(For $\gamma = \beta$ the eigenvectors do not depend on α, β, γ.)

Applying Lemma 2 either directly to A, or to G with $N = m-1$, $\alpha = 2$ and $\beta = \gamma = -1$, yields

$$\mu_k^G = 2 - 2\cos\frac{k\pi}{m} = 4\sin^2\left(\frac{k\pi}{2m}\right)$$

$$\mu_k^A = 1 - 4\lambda \sin^2 \frac{k\pi}{2m}\,.$$

By Lemma 1:

$$\begin{aligned} \text{stability} &\iff \left| 1 - 4\lambda \sin^2 \frac{k\pi}{2m} \right| < 1, \quad k = 1, ..., m-1 \\ &\iff \lambda > 0 \text{ and } -1 < 1 - 4\lambda \sin^2 \frac{k\pi}{2m}, \quad \text{or } \frac{1}{2} > \lambda \sin^2 \frac{k\pi}{2m} \end{aligned}$$

For the second inequality $\lambda \le 1/2$ is sufficient. In summary,

For $0 < \lambda \le \frac{1}{2}$ the explicit method $w^{(\nu+1)} = Aw^{(\nu)}$ with $A = A_{\text{expl}}$ is stable.

Because of $\lambda = \frac{\Delta\tau}{\Delta x^2}$, or $0 < \Delta\tau \le \frac{1}{2}\Delta x^2$, the step sizes $\Delta\tau$ and Δx can not be chosen independent of each other.

Conclusion: This explicit method is not satisfying.

E. An Implicit Method

The backward difference quotient

$$\frac{\partial y_{i,\nu}}{\partial \tau} = \frac{y_{i,\nu} - y_{i,\nu-1}}{\Delta\tau} + O(\Delta\tau).$$

leads to

$$-\lambda w_{i+1,\nu} + (1+2\lambda) w_{i,\nu} - \lambda w_{i-1,\nu} = w_{i,\nu-1}\,.$$

This is a system of coupled linear equations for the $w_{i,\nu}$. With

$$A := A_{\text{impl}} := \begin{pmatrix} 1+2\lambda & -\lambda & & 0 \\ -\lambda & \ddots & \ddots & \\ & \ddots & \ddots & -\lambda \\ 0 & & -\lambda & 1+2\lambda \end{pmatrix}$$

the vector $w^{(\nu)}$ is defined implicitly as solution of $Aw^{(\nu)} = w^{(\nu-1)}$, or

$$Aw^{(\nu+1)} = w^{(\nu)}\,,\ \nu = 0, \ldots, \nu_{\max} - 1.$$

(We still use the preliminary boundary conditions $w_{0,\nu} = w_{m,\nu} = 0$.) This method is called *backward difference method* or *backward time centered space (BTCS)* or *fully implicit*.

Stability

The above Lemmas imply that the method is *unconditionally stable*, and $\Delta\tau$ and Δx can be chosen independent of each other.

Costs

$\nu_{\max} O(m)$, since only one LR-decomposition of A for $\nu = 0$ is necessary (cheap for a tridiagonal matrix). For each ν, only one backward loop is required, which costs $O(m)$ operations.

A weakness of this method (and of the explicit method) is the accuracy of the first order in $\Delta\tau$, the error is of the order

$$O(\Delta x^2) + O(\Delta\tau)\,.$$

4.3 Crank–Nicolson Method

It would be desirable to have a stable method with error $O(\Delta\tau^2)$ for

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

From the previous section, we assemble the forward quotient for ν

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}$$

and the backward quotient for $\nu + 1$

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}}{\Delta x^2}\,.$$

Adding both equations yields the scheme

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{1}{2\Delta x^2}\left(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu} + w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}\right).$$

Theorem (Crank–Nicolson)

For this scheme the following assertions hold:

1.) For $y \in C^4$ the method is of the order $O(\Delta x^2) + O(\Delta\tau^2)$.

2.) For each ν a system of linear equations in tridiagonal form must be solved.

3.) The method is stable for all $\Delta\tau > 0$.

Proof:

1.) With the notation

$$\delta_{xx} w_{i,\nu} := \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}$$

the Taylor expansion for $y \in C^4$ yields

$$\delta_{xx} y_{i,\nu} = \frac{\partial^2}{\partial x^2} y_{i,\nu} + \frac{\Delta x^2}{12} \frac{\partial^4}{\partial x^4} y_{i,\nu} + O(\Delta x^4)\,.$$

Then the local discretization error

$$\varepsilon := \frac{y_{i,\nu+1} - y_{i,\nu}}{\Delta\tau} - \frac{1}{2}\left(\delta_{xx} y_{i,\nu} + \delta_{xx} y_{i,\nu+1}\right)$$

satisfies

$$\varepsilon = O(\Delta x^2) + O(\Delta\tau^2)\,.$$

2.) With $\lambda := \frac{\Delta\tau}{\Delta x^2}$ the scheme is rewritten

$$-\frac{\lambda}{2} w_{i-1,\nu+1} + (1+\lambda) w_{i,\nu+1} - \frac{\lambda}{2} w_{i+1,\nu+1}$$

$$= \frac{\lambda}{2} w_{i-1,\nu} + (1-\lambda) w_{i,\nu} + \frac{\lambda}{2} w_{i+1,\nu}\,.$$

With the preliminary boundary conditions $w_{0,\nu} = w_{m,\nu} = 0$ this is the matrix-vector system

$$A w^{(\nu+1)} = B w^{(\nu)}$$

where

$$A := \begin{pmatrix} 1+\lambda & -\frac{\lambda}{2} & & 0 \\ -\frac{\lambda}{2} & \ddots & \ddots & \\ & \ddots & \ddots & -\frac{\lambda}{2} \\ 0 & & -\frac{\lambda}{2} & 1+\lambda \end{pmatrix}, \quad B := \begin{pmatrix} 1-\lambda & \frac{\lambda}{2} & & 0 \\ \frac{\lambda}{2} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{\lambda}{2} \\ 0 & & \frac{\lambda}{2} & 1-\lambda \end{pmatrix}.$$

By the theorem of Gerschgorin the eigenvalues of A lie between 1 and $1+2\lambda$. So, zero is no eigenvalue, A is non-singular, and the system of equations has a unique solution.

3.) Set

$$A = I + \tfrac{\lambda}{2} G \;\text{ with }\; G := \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix} \;\text{ and }\; B = I - \tfrac{\lambda}{2} G\,.$$

Then

$$\underbrace{(2I + \lambda G)}_{=:C} w^{(\nu+1)} = (2I - \lambda G) w^{(\nu)}$$

$$= (4I - 2I - \lambda G) w^{(\nu)}$$

$$= (4I - C) w^{(\nu)}\,,$$

which leads to the explicit form

$$w^{(\nu+1)} = (4C^{-1} - I)w^{(\nu)}.$$

By Lemma 1 the stability requirement is

$$\left| \frac{4}{\mu_k^C} - 1 \right| < 1 \quad \text{for all } k.$$

By Section 4.2D, the eigenvalues μ_k^C of C are

$$\mu_k^C = 2 + \lambda \mu_k^G = 2 + 4\lambda \sin^2 \frac{k\pi}{2m} \; > 2\,.$$

Hence the method is stable for all $\lambda > 0$ ($\Delta\tau > 0$).

Algorithm (Crank–Nicolson)

start: Choose m, $\nu_{\max}$; calculate $\Delta x, \Delta\tau$
$w_i^{(0)} = y(x_i, 0) \quad (0 \le i \le m)$
LR-decomposition (or RL-decomposition) of A
loop: for $\nu = 0, 1, ..., \nu_{\max} - 1$:
$c := Bw^{(\nu)} + 0$ (preliminary boundary conditions 0)
Solve $Ax = c$ (using the LR/RL-decomposition)
$w^{(\nu+1)} := x$

4.4 Boundary Conditions

Since the initial conditions are active only for the interval $x_{\min} \le x \le x_{\max}$, boundary conditions

$y(x, \tau)$ for $x_{\min}$ and $x_{\max}$, or
$w_{0,\nu}$ and $w_{m,\nu}$ for $\nu = 1, ..., \nu_{\max}$

are needed to make the problem well-posed. These boundary conditions are artificial and have an influence on the accuracy of numerical solutions.

In the S-world, the GBM assumption for S_t implies

$$\begin{aligned} S(0) = 0 \quad &\Rightarrow \quad S_t = 0 \text{ for all } t \\ S(0) \to \infty \quad &\Rightarrow \quad S_t \text{ large for all } t \le T. \end{aligned}$$

Hence for all t

$$\begin{aligned} V_{\mathrm{C}}(S,t) &= 0 \quad \text{for } S = 0, \quad \text{and} \\ V_{\mathrm{P}}(S,t) &\to 0 \quad \text{for } S \to \infty\,. \end{aligned}$$

We use these boundary conditions as approximations also for $S_{\min} \approx 0$ and $S_{\max}$ large. And for the transformed x-values, provided $-x_{\min} = -x_0$ and $x_{\max} = x_m$ are large enough, as well:

$$\begin{array}{lll} \text{call: “left”} & (S=0) & V=0, \quad \text{i.e. } w_{0,\nu}=0 \text{ for all } \nu \\ \text{put: “right”} & (S\to\infty) & V=0, \quad \text{i.e. } w_{m,\nu}=0 \text{ for all } \nu \end{array}$$

So far, the argumentation holds for both European and American options. Now we turn to the question, what are reasonable V-values on the other ends of the interval?

We begin with **European** options, and apply the *put-call parity*

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

This yields the boundary conditions at the other ends:

$$\begin{aligned} V_{\rm C}(S,t) &= S\mathrm{e}^{-\delta(T-t)} - K\mathrm{e}^{-r(T-t)} \quad \text{for } S\to\infty \\ V_{\rm P}(S,t) &= K\mathrm{e}^{-r(T-t)} - S\mathrm{e}^{-\delta(T-t)} \quad \text{for } S\to 0\,. \end{aligned}$$

After the transformation $(S,t)\to(x,\tau)$ the asymptotic behavior is:

$$\begin{array}{l} y(x,\tau) = r_1(x,\tau) \text{ for } x\to-\infty, \quad y(x,\tau) = r_2(x,\tau) \text{ for } x\to\infty, \text{ with} \\ \text{call:} \quad r_1(x,\tau) := 0, \quad r_2(x,\tau) := \exp\left(\frac{1}{2}(q_\delta+1)x + \frac{1}{4}(q_\delta+1)^2\tau\right) \\ \text{put:} \quad r_1(x,\tau) := \exp\left(\frac{1}{2}(q_\delta-1)x + \frac{1}{4}(q_\delta-1)^2\tau\right), \quad r_2(x,\tau) := 0 \end{array}$$

For the finite interval $a := x_{\min} \le x \le x_{\max} =: b$, r_1, r_2 are the dominating terms of the boundary conditions and can be used as approximations. Accordingly, we choose

$$\begin{aligned} w_{0,\nu} &= r_1(a,\tau_\nu) \\ w_{m,\nu} &= r_2(b,\tau_\nu)\,. \end{aligned}$$

These boundary conditions are of Dirichlet-type. In the Crank–Nicolson scheme the boundary conditions lead to the additional terms

$$\begin{aligned} -\frac{\lambda}{2}\,w_{0,\nu+1} &= -\frac{\lambda}{2}\,r_1(a,\tau_{\nu+1}) \\ \frac{\lambda}{2}\,w_{0,\nu} &= \frac{\lambda}{2}\,r_1(a,\tau_\nu) \end{aligned}$$

and

$$\begin{aligned} -\frac{\lambda}{2}\,w_{m,\nu+1} &= -\frac{\lambda}{2}\,r_2(b,\tau_{\nu+1}) \\ \frac{\lambda}{2}\,w_{m,\nu} &= \frac{\lambda}{2}\,r_2(b,\tau_\nu), \end{aligned}$$

which are represented by the vector

$$d^{(\nu)} := \frac{\lambda}{2}\cdot\begin{pmatrix} r_1(a,\tau_{\nu+1}) + r_1(a,\tau_\nu) \\ 0 \\ \vdots \\ 0 \\ r_2(b,\tau_{\nu+1}) + r_2(b,\tau_\nu) \end{pmatrix} = \frac{\lambda}{2}\cdot\begin{pmatrix} w_{0,\nu} + w_{0,\nu+1} \\ 0 \\ \vdots \\ 0 \\ w_{m,\nu} + w_{m,\nu+1} \end{pmatrix}$$

In the Crank–Nicolson algorithm the right-hand side of the system of equations is now

$$c := Bw^{(\nu)} + d^{(\nu)} ;$$

the matrix A is unchanged.

We still need to specify the boundary conditions (b.c.) for American-style options on the "other" side, namely, left-hand b.c. for the put and right-hand b.c. for the call.

4.5 Early-Exercise Structure — Free-Boundary Problems

For a better understanding of American options a further analysis of the solution structure is helpful.

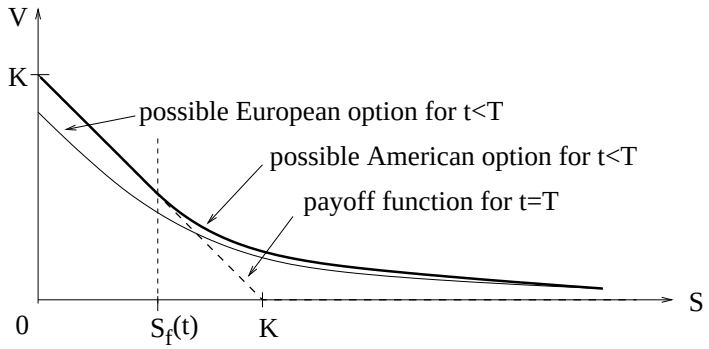
Basics (for the put under the Black–Scholes model)

a) $V(S,t)$ is continuous, $V \geq 0$, and $V \to 0$ for $S \to \infty$.

b) For $r > 0$ and all $t < T$

$$V_{\rm P}^{\rm Eu}(0,t) = K{\rm e}^{-r(T-t)} < K\,,$$

which implies: There is a $S^* = S^{*\rm Eu}(t)$ such that $V_{\rm P}^{\rm Eu}(S^*,t) = K - S^*$.



c) $V_{\rm P}$ is monotonic decreasing w.r.t. S, and convex. [R.C. Merton: Theory of Rational Option Pricing (1973)]

Hence S^* is unique.

d) $V_{\rm P}^{\rm Am} \geq (K-S)^+$ and $V_{\rm P}^{\rm Am} \leq K$.

Assertion 1

Also for the American put with $r > 0$ there is an $S^* > 0$, such that $V_{\rm P}^{\rm Am}(S^*,t) = K - S^*$.

Proof:

Assume: $V_{\rm P}^{\rm Am} > K - S$ for all $S > 0$.
Then

$$-V_{\rm P}^{\rm Am} + K - S < 0\,,$$

i.e., exercising the put leads to a loss for all S. Accordingly, early exercise does not make sense, and hence $V_{\rm P}^{\rm Am} = V_{\rm P}^{\rm Eu}$. Consequently,

$$V_{\rm P}^{\rm Eu} = V_{\rm P}^{\rm Am} > K - S\,,$$

for all $S > 0$, which contradicts the implication of b).

Redefine $S^* := \max\{\, S \mid V^{\mathrm{Am}}(S,t) = K - S \,.\}$

By c) and d), $V = K - S$ for all $S < S^*$.

It remains to investigate the behavior of V for $S \geq S^*$.

Assertion 2

The right-hand side derivative of the function $V_{\mathrm{P}}^{\mathrm{Am}}(S,t)$ at S^* has the value -1.

Proof (outline):

We use the notation $V := V_{\mathrm{P}}^{\mathrm{Am}}$, and $\frac{\partial V}{\partial S}$ stands for the right-hand derivative.
For the right-hand derivative $\frac{\partial V(S^*,t)}{\partial S} < -1$ is impossible, because otherwise for $S > S^*$ the property d) is violated. Hence $\frac{\partial V(S^*,t)}{\partial S} \geq -1$.
Assumption:

$$\frac{\partial V(S^*,t)}{\partial S} > -1\,.$$

We lead this to a contradiction as follows:
Build a portfolio: $\Pi := V + S$, with initial wealth

$$\Pi^* := V + S^*\,.$$

$(= K$; borrow from the bank the amount $K.)$ For our GBM

$$\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,\mathrm{d}W$$

assume that $\mathrm{d}t$ is so small that $\sqrt{\mathrm{d}t} \gg \mathrm{d}t$. If $\mathrm{d}t$ is small enough, then (intuitively)

$$\mathrm{d}W > 0 \quad \Longleftrightarrow \quad \mathrm{d}S > 0\,.$$

The Itô-Lemma leads to

$$\begin{aligned}\mathrm{d}\Pi &= (\ldots)\mathrm{d}t + \frac{\partial \Pi}{\partial S}\sigma S\,\mathrm{d}W \\ &= O(\mathrm{d}t) + \left(\frac{\partial V}{\partial S} + 1\right)\sigma S\,\mathrm{d}W.\end{aligned}$$

For $\mathrm{d}S > 0$ this is positive for sufficiently small $\mathrm{d}t$ because $\mathrm{d}W > 0$.
For $\mathrm{d}S < 0$ the wealth of the portfolio is $\Pi \equiv K$ and hence $\mathrm{d}\Pi = 0$.
In summary: $\mathsf{E}(\mathrm{d}\Pi) > 0$, and $\mathsf{E}(\mathrm{d}\Pi)$ is of the order $O(\sqrt{\mathrm{d}t})$. Sell the portfolio after $\mathrm{d}t$ and expect the following balance

$$\begin{aligned}\Pi^* + \mathsf{E}(\mathrm{d}\Pi) - K\mathrm{e}^{r\mathrm{d}t} &= V + S^* + \mathsf{E}(\mathrm{d}\Pi) - K(1 + O(\mathrm{d}t)) \\ &= \mathsf{E}(\mathrm{d}\Pi) + O(\mathrm{d}t)\end{aligned}$$

This is positive because $\mathsf{E}(\mathrm{d}\Pi)$ dominates, hence an arbitrage, and we have arrived at a contradiction to the no-arbitrage principle.

Often related proofs use an argument of maximizing the value of the option. In this way, the *perpetual option* (an option that does not expire) can be analyzed, see the exercises.

In summary, for an American put the following holds:

(1) $V(S^*,t) = K - S^*$ (Dirichlet b.c.)
(2) $\frac{\partial V(S^*,t)}{\partial S} = -1$ (Neumann b.c.)
(3) $V(S,t) > K - S$ for $S > S^*$
(4) $V(S,t) = K - S$ for $S \le S^*$

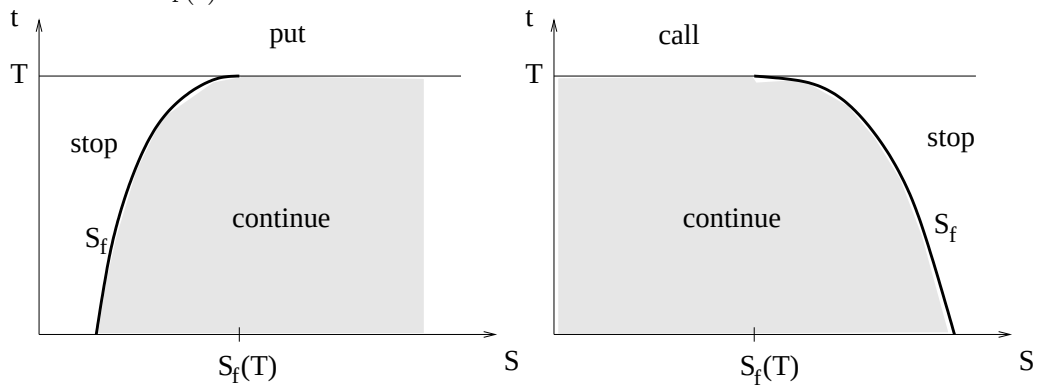
The property (2) is called *high contact*, or *smooth pasting*.

Such an S^* exists for each t. This defines a function, which we denote $S_{\rm f}(t)$. The "f" stands for *free boundary*.

The curve $S_{\rm f}(t)$ cuts the half strip into two parts, namely,

1.) $S > S_{\rm f}$, called *continuation region* of the put.
2.) $S \le S_{\rm f}$, called *stopping region* of the put.

For standard options without discrete dividend payments, these domains are simply connected. The curve $S_{\rm f}(t)$ is the interface.



Early-Exercise Curve

The curve $S_{\rm f}(t)$ is the *early-exercise curve* by the following reasons:

1.) In case $V_{\rm P} > (K-S)^+$, exercising amounts to $-V + K - S < 0$. This is a loss. Consequently, the holder continues to *hold* the option.

2.) In case the price S passes the curve, $S < S_{\rm f}(t)$, then immediate exercising makes sense ("stopping"), because the amount K can be invested, leading for $r > 0$, $t < T$ to the profit:

$$K{\rm e}^{r(T-t)} - K = K({\rm e}^{r(T-t)} - 1).$$

By exercising, the final balance of $K{\rm e}^{r(T-t)}$ is larger than $S{\rm e}^{\delta(T-t)}$, at least for $r(T-t) < 1$.

Free-Boundary Problem means:

The Black–Scholes equation is valid only in the continuation region, not in the stopping region. Hence the domain for the BS equation for an American-style put is

$$S_{\rm f}(t) < S < \infty\,.$$

The left-hand boundary $S_{\rm f}(t)$ is "free" in the sense that it is unknown initially. It is calculated numerically, based on the additional boundary condition provided by the contact condition $\frac{\partial V}{\partial S} = -1$. This condition fixes the location of $S_{\rm f}(t)$.

The properties of a call are derived analogously.

Important Properties of the early-exercise curve in case **of a put** under the Black–Scholes model (continuous dividend rate $\delta \geq 0$ possible, but discrete dividend here excluded!) are:

1.) $S_{\rm f}(t)$ is continuously differentiable for $t < T$.

2.) $S_{\rm f}(t)$ is monotonic increasing.

3.)
$$S_{\rm f}(T) := \lim_{\substack{t \to T \\ t<T}} S_{\rm f}(t) = \begin{cases} K & \text{für } 0 \leq \delta \leq r \\ \frac{r}{\delta} K & \text{für } r < \delta \end{cases}$$

Proof of 3.) (notation $V = V_{\rm P}^{\rm Am}$)

$V^{\rm Am} \geq V^{\rm Eu}$ implies $S_{\rm f}(t) \leq S^{*\rm eu}(t) < K$ for all $t < T$. Hence $S_{\rm f}(T) \leq K$.

Notice that $V \geq \Psi$ implies
$$\frac{\partial V(S,t)}{\partial t} \leq 0$$
for $t = T$. To prepare for some indirect proofs, we first study for $S_{\rm f}(T) < K$ how the BS equation is consistent with the sign of $\frac{\partial V}{\partial t}$. At $t = T$ and for $S_{\rm f}(T) < S < K$
$$V(S,T) = K - S$$
holds, and the BS equation is
$$\frac{\partial V}{\partial t} + 0 - (r-\delta)S - rV = 0$$
$$\Longrightarrow \quad \frac{\partial V(S,T)}{\partial t} = rK - \delta S\,.$$

We will check for which combinations of (δ, r) the sign of $\frac{\partial V(S,T)}{\partial t}$ is consistent with $\frac{\partial V}{\partial t} \leq 0$.

Case $\delta > r$

Here $\frac{r}{\delta}K < K$. Then either $S_{\rm f}(T) = \frac{r}{\delta}K$ (the assertion), or there exists one of two open intervals (i) $S_{\rm f}(T) < \frac{r}{\delta}K$ and (ii) $\frac{r}{\delta}K < S_{\rm f}(T)$.

(i) For S in the interval $S_{\rm f}(T) < S < \frac{r}{\delta}K$ we have $\frac{\partial V(S,T)}{\partial t} = rK - \delta S > 0$, a contradiction to
$$\frac{\partial V}{\partial t} \leq 0\,.$$

(ii) For each S in the interval $\frac{r}{\delta}K < S < S_{\rm f}(T)$ there is a small $\mathrm{d}t$ such that $(S, T-\mathrm{d}t)$ is in the stopping region. The inequality $rK < \delta S$ holds, and thus $rK\mathrm{d}t < \delta S \mathrm{d}t$, which leads to
$$K(\mathrm{e}^{r\mathrm{d}t} - 1) < S(\mathrm{e}^{\delta \mathrm{d}t} - 1)\,.$$

Since $S < K$, this means that in case of stopping of the option for the time interval $\mathrm{d}t$ the dividend yield is larger than the return of investing K at the risk-free rate. Hence early exercise is not optimal, which is in conflict with the meaning of $S < S_{\mathrm{f}}(t)$.

Hence $S_{\mathrm{f}}(T) = \frac{r}{\delta}K$ holds in case $\delta > r$.

Case $\delta \le r$

Assume $S_{\mathrm{f}}(T) < K$. Then for S in the interval $S_{\mathrm{f}}(T) < S < K$ a contradiction is obtained from

$$\underbrace{\frac{\partial V}{\partial t}}_{\le 0} = \underbrace{rK - \delta S}_{>0} .$$

For a call with $\delta > 0$ but without discrete dividend:

1.) $S_{\mathrm{f}}(t)$ is continuously differentiable for $t < T$.

2.) $S_{\mathrm{f}}(t)$ is monotonic decreasing.

3.)

$$S_{\mathrm{f}}(T) := \lim_{\substack{t \to T \\ t<T}} S_{\mathrm{f}}(t) = \max\left\{K, \frac{r}{\delta}K\right\}$$

Remark: In case of discrete dividend payment the above assertions must be modified. In particular, S_{f} then is not continuous! For example, for an American put early exercise is not optimal within a certain time interval before ex-dividend date.

4.6 Linear Complementarity

We need a numerical method that does not use the unknown S_{f} explicitly.

A. Inequality

As argued above, for American-style options the Black–Scholes equation does not hold in the entire half strip $0 \le t \le T,\ 0 < S$, but only in the continuation region. Now, what happens in the stopping region?

With the notation:

$$\mathcal{L}_{\mathrm{BS}}(V) := \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-\delta)S\frac{\partial V}{\partial S} - rV$$

the Black–Scholes equation can be written

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\mathrm{BS}}(V) = 0 .$$

For for a put and $S < S_{\mathrm{f}}$ the identity $V \equiv \Psi$ holds, so

$$V = K - S, \quad \frac{\partial V}{\partial S} = -1, \quad \frac{\partial^2 V}{\partial S^2} = 0, \quad \frac{\partial V}{\partial t} = 0,$$

hence

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) = \delta S - rK\,.$$

For $S < S_{\rm f}(T)$ we conclude

in case $r < \delta$: $S < \frac{r}{\delta}K$, hence $\delta S - rK < 0$

in case $r \geq \delta$: $\delta S - rK \leq r(S-K) < 0$ because $S < K$

and thus

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) < 0\,.$$

For a call the same result follows by analogous arguments.

In summary for all $S > 0,\ 0 \leq t \leq T$ the partial **differential inequality**

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) \leq 0$$

holds, both for put and call.

Overview

put: $V_{\rm P}^{\rm Am} = K - S$ for $S \leq S_{\rm f}$

$V_{\rm P}^{\rm Am}$ solves the BS equation for $S > S_{\rm f}$

contact condition: $\dfrac{\partial V(S_{\rm f},t)}{\partial S} = -1$

call: $V_{\rm C}^{\rm Am} = S - K$ for $S \geq S_{\rm f}$

$V_{\rm C}^{\rm Am}$ solves the BS equation for $S < S_{\rm f}$

contact condition: $\dfrac{\partial V(S_{\rm f},t)}{\partial S} = 1$

The second derivative of V with respect to S is not continuous at $S_{\rm f}$. That is, the value function V is smooth in the interior of the continuation region, but not on the entire half strip.

Remark: The transformation of Section 4.1 leads to

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) = -\frac{\partial y}{\partial \tau} + \frac{\partial^2 y}{\partial x^2}\,.$$

B. Formulation with Penalty Term

A unified treatment of $\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) \leq 0$ on the entire half strip is possible. To this end, introduce a suitable function $p(V) \geq 0$ requiring the penalty PDE

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\rm BS}(V) + p(V) = 0$$

to hold. The *penalty term* p should be 0 in the continuation region, and positive in the stopping region. The distance to $S_{\rm f}$ is not known, but the distance $V - \Psi$ of V to the payoff Ψ is available and is used as control. One example of a penalty function is

$$p(V) := \frac{\epsilon}{V - \Psi} \quad \text{for a small } \epsilon > 0\,.$$

Let $V_\epsilon(S,t)$ denote the solution of the penalty PDE. Two extreme cases characterize the effect of the penalty term for (S,t) in the continuation area and in the stopping area:

- $V_\epsilon - \Psi \gg \epsilon$ implies $p \approx 0$. Then essentially the Black–Scholes equation results.
- $0 < V_\epsilon - \Psi \ll \epsilon$ implies a large value of p, which means that the BS-part of the equation is dominated by p. The BS equation is switched off, and $V_\epsilon \approx \Psi$.

The corresponding branches of the solution V_ϵ may be called the "continuation branch" ($p \approx 0$) and the "stopping branch" ($V_\epsilon \approx \Psi$). Obviously these two branches approximate the true solution V of the Black–Scholes problem. The intermediate range $V_\epsilon - \Psi \approx O(\epsilon)$ characterizes a boundary layer between the continuation branch and the stopping branch. In this layer around the early-exercise curve S_f the solution V_ϵ can be seen as a connection between the BS surface and the payoff plane.[a]

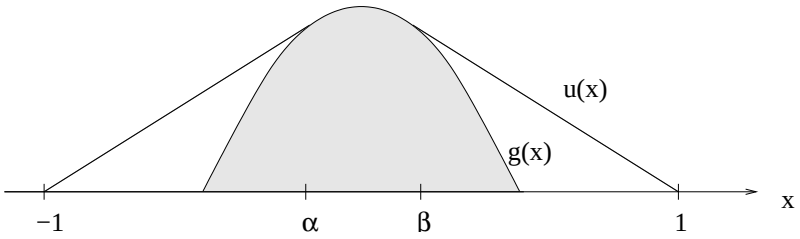
Remarks: p and the resulting PDE are nonlinear in V. An implementation that avoids $V_\epsilon \le \Psi$ is not easy; not every choice of ϵ or Δt will be successful.

Penalty methods are powerful in general. But for the relatively simple situation of the single-asset American option, a more elegant solution is possible. We shall describe this approach next.

C. Simple Obstacle Problem

Consider an "obstacle" g satisfying[b]

$$g > 0 \text{ for a subinterval of } -1 < x < 1,\ g \in C^2,\ g'' \le 0,\ g(-1) < 0 \text{ and } g(1) < 0.$$



A function $u \in C^1$ with minimal length, and with $u(-1) = u(1) = 0$ and $u \ge g$ for the g of the figure can be characterized as follows: There is α, β such that

$$\begin{aligned} -1 < x < \alpha: &\quad u'' = 0 \quad (u > g) \\ \alpha < x < \beta: &\quad u = g \quad (u'' = g'' \le 0) \\ \beta < x < 1: &\quad u'' = 0 \quad (u > g) \end{aligned}$$

This expresses a complementarity in the sense of

$$\begin{aligned} &\text{in case } u - g > 0, \text{ then } u'' = 0, \\ &\text{in case } u - g = 0, \text{ then } u'' \le 0. \end{aligned}$$

[a] This is illustrated in Topic 9, see the *Topics for CF* on the homepage `www.compfin.de`; Topic 9 also also illustrates the penalty function p.

[b] This and other parts of this chapter were inspired by [P. Wilmott, J. Dewynne, S. Howison: The Mathematics of Financial Derivatives].

For options, the analogy is

$$\begin{aligned} &\text{in case } V-\Psi>0, \text{ then } \dot{V}+\mathcal{L}_{\mathrm{BS}}(V)=0\,,\\ &\text{in case } V-\Psi=0, \text{ then } \dot{V}+\mathcal{L}_{\mathrm{BS}}(V)\le 0\,. \end{aligned}$$

For the simple obstacle problem an equivalent formulation is

$$\begin{aligned} &\text{Find a function } u \text{ such that}\\ &u''(u-g)=0, \quad -u''\ge 0, \quad u-g\ge 0, \qquad\qquad (*)\\ &u(-1)=u(1)=0,\ u\in\mathcal{C}^1[-1,1]. \end{aligned}$$

In this version, α and β do not occur explicitly. After the numerical solution, the values of α and β will become apparent. The problem $(*)$ is a **linear complementarity problem** (LCP); it has the form

$$\mathcal{A}\cdot\mathcal{B}=0,\ \mathcal{A}\ge 0,\ \mathcal{B}\ge 0, \qquad \text{for suitable } \mathcal{A},\mathcal{B}.$$

For a **numerical solution** of the simple obstacle problem we use the grid

$$x_i=-1+i\Delta x, \quad \Delta x:=\frac{2}{m}, \quad g_i:=g(x_i)\,.$$

This leads to the discrete form of the obstacle problem

$$\left.\begin{aligned} (w_{i-1}-2w_i+w_{i+1})(w_i-g_i)&=0,\\ -w_{i-1}+2w_i-w_{i+1}&\ge 0,\\ w_i-g_i&\ge 0 \end{aligned}\right\} \quad \text{for } 0<i<m$$

with $w_0=w_m=0$. Correspondingly we define

$$G:=\begin{pmatrix} 2 & -1 & & 0\\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1\\ 0 & & -1 & 2 \end{pmatrix} \quad \text{and} \quad w:=\begin{pmatrix} w_1\\ \vdots\\ w_{m-1}\end{pmatrix},\ g:=\begin{pmatrix} g_1\\ \vdots\\ g_{m-1}\end{pmatrix}$$

to rewrite the linear complementarity problem in vector notation:

$$\left.\begin{aligned} (w-g)^{tr}Gw&=0\\ Gw&\ge 0\\ w-g&\ge 0 \end{aligned}\right\}$$

D. Complementarity of the Black–Scholes Problem

Specifically for the American put let us formulate the transformed problem

$$\frac{\partial y}{\partial\tau}=\frac{\partial^2 y}{\partial x^2} \quad \text{as long as} \quad V_{\mathrm{P}}^{\mathrm{Am}}>(K-S)^+$$

as LCP. This requires to transform also the side conditions:

$$\begin{aligned} V_{\mathrm{P}}^{\mathrm{Am}}(S,t) &\geq (K-S)^+ \\ &= K\max\{1-\mathrm{e}^x, 0\} \end{aligned}$$

with $q = \frac{2r}{\sigma^2}$ leads to (here specially for $\delta = 0$):

$$\begin{aligned} y(x,\tau) &\geq \exp\{\tfrac{1}{2}(q-1)x + \tfrac{1}{4}(q+1)^2\tau\}\max\{1-\mathrm{e}^x,\, 0\} \\ &= \exp\{\tfrac{1}{4}(q+1)^2\tau\}\max\{(1-\mathrm{e}^x)\mathrm{e}^{\frac{1}{2}(q-1)x},\, 0\} \\ &= \exp\{\tfrac{1}{4}(q+1)^2\tau\}\max\{\mathrm{e}^{\frac{1}{2}(q-1)x} - \mathrm{e}^{\frac{1}{2}(q+1)x},\, 0\} \\ &=: g(x,\tau). \end{aligned}$$

It turns out that both the boundary- and the initial conditions can be written with this g:

$$\begin{aligned} y(x,0) &= g(x,0) \\ y(x_{\min},\tau) &= g(x_{\min},\tau) \\ y(x_{\max},\tau) &= g(x_{\max},\tau) \end{aligned}$$

Now the linear complementarity problem for the American put is written

$$\left.\begin{aligned} \left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2}\right)(y-g) &= 0 \\ \frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} &\geq 0 \\ y - g &\geq 0 \end{aligned}\right\}$$

with boundary- and initial conditions as stated above.

This formulation is identical to that one for the general case $\delta \neq 0$, except for an adapted g:

$$q = \frac{2r}{\sigma^2}\,, \quad q_\delta = \frac{2(r-\delta)}{\sigma^2}\,,$$

put: $(r > 0)$ $\quad g(x,\tau) := \exp\{\tfrac{1}{4}((q_\delta-1)^2 + 4q)\tau\}\max\{\mathrm{e}^{\frac{1}{2}(q_\delta-1)x} - \mathrm{e}^{\frac{1}{2}(q_\delta+1)x},\, 0\}$

call: $(\delta > 0)$ $\quad g(x,\tau) := \exp\{\tfrac{1}{4}((q_\delta-1)^2 + 4q)\tau\}\max\{\mathrm{e}^{\frac{1}{2}(q_\delta+1)x} - \mathrm{e}^{\frac{1}{2}(q_\delta-1)x},\, 0\}$

Remark: For an American call without dividend and $r > 0$, $t < T$, the equality $V_{\mathrm{C}}^{\mathrm{Am}} = V_{\mathrm{C}}^{\mathrm{Eur}}$ holds. *Proof* for $\delta = 0$:

$$V_{\mathrm{C}}^{\mathrm{Am}} \geq V_{\mathrm{C}}^{\mathrm{Eur}} \geq S - K\mathrm{e}^{-r(T-t)} > S - K$$

Hence $V_{\mathrm{C}}^{\mathrm{Am}}$ >payoff; no early exercise.

4.7 Numerical Realization

A. Discretization

Now we discretize the LCP-problem with the grid from Section 4.2B:

$$\frac{w_{i,\nu+1}-w_{i,\nu}}{\Delta\tau} = \theta\frac{w_{i+1,\nu+1}-2w_{i,\nu+1}+w_{i-1,\nu+1}}{\Delta x^2} + (1-\theta)\frac{w_{i+1,\nu}-2w_{i,\nu}+w_{i-1,\nu}}{\Delta x^2}$$

("theta method"), with $\theta = 0$ (explicit method), $\theta = \frac{1}{2}$ (Crank–Nicolson) or $\theta = 1$ (BTCS-method). With the notation $\lambda := \frac{\Delta\tau}{(\Delta x)^2}$ the inequality $\frac{\partial y}{\partial\tau} - \frac{\partial^2 y}{\partial x^2} \geq 0$ becomes

$$\begin{aligned} w_{i,\nu+1} - \lambda\theta(w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}) \\ - w_{i,\nu} - \lambda(1-\theta)(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}) \geq 0\,. \end{aligned} \qquad (*)$$

Ordering these terms leads to define for the ν-level terms

$$b_{i,\nu} := w_{i,\nu} + \lambda(1-\theta)(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}),\ \text{for } i = 2,\ldots,m-2.$$

The boundary conditions are included in $b_{1,\nu}$ and $b_{m-1,\nu}$:

$$\begin{aligned} b_{1,\nu} &:= w_{1,\nu} + \lambda(1-\theta)(w_{2,\nu} - 2w_{1,\nu} + g_{0,\nu}) + \lambda\theta g_{0,\nu+1} \\ b_{m-1,\nu} &:= w_{m-1,\nu} + \lambda(1-\theta)(g_{m,\nu} - 2w_{m-1,\nu} + w_{m-2,\nu}) + \lambda\theta g_{m,\nu+1}\,, \end{aligned}$$

where

$$g_{i,\nu} := g(x_i,\tau_\nu) \quad (0 \leq i \leq m,\ 0 \leq \nu \leq \nu_{\max}).$$

This completes the vector

$$b^{(\nu)} := (b_{1,\nu},\ldots,b_{m-1,\nu})^{tr}$$

and analogously, $w^{(\nu)}, g^{(\nu)}$. With the matrix

$$A := \begin{pmatrix} 1+2\lambda\theta & -\lambda\theta & & 0 \\ -\lambda\theta & \ddots & \ddots & \\ & \ddots & \ddots & -\lambda\theta \\ 0 & & -\lambda\theta & 1+2\lambda\theta \end{pmatrix} \in \mathbb{R}^{(m-1)\times(m-1)}$$

the problem $(*)$ is reformulated as

$$Aw^{(\nu+1)} \geq b^{(\nu)} \quad \text{for all } \nu.$$

And $y - g \geq 0$ reads

$$w^{(\nu)} \geq g^{(\nu)},$$

and finally $\left(\frac{\partial y}{\partial\tau} - \frac{\partial^2 y}{\partial x^2}\right)(y-g) = 0$ becomes

$$\left(Aw^{(\nu+1)} - b^{(\nu)}\right)^{tr}\left(w^{(\nu+1)} - g^{(\nu+1)}\right) = 0.$$

This constitutes the following macro-algorithm:

Algorithm

For $\nu = 0, 1, ..., \nu_{\max} - 1$:
compute $g := g^{(\nu+1)}$, $b := b^{(\nu)}$, as above;
compute w as solution of
$$Aw - b \geq 0, \quad w \geq g, \quad (Aw - b)^{tr}(w - g) = 0. \qquad (**)$$
set $w^{(\nu+1)} := w$

For each time level ν the LCP $(**)$ must be solved. This topic is analyzed next.

Assertion

With the transformation $x := w - g$ and $y := Aw - b$ the LCP $(**)$ is equivalent to

Compute vectors x, y, such that for $\hat{b} := b - Ag$ the following holds:
$$Ax - y = \hat{b}, \quad x \geq 0, \quad y \geq 0, \quad x^{tr} y = 0 \qquad (***)$$

Notice that in this context x, y are general vectors $\in \mathbb{R}^{m-1}$.

Proof: Apply the transformation.

Lemma

The problem $(***)$ has a unique solution.

Proof:

1.) Define
$$G(x) := \frac{1}{2} x^{tr} A x - \hat{b}^{tr} x$$
with the A from above,
$$A = \mathrm{I} + \lambda\theta \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}$$
By Lemma 2 of Section 4.2D the eigenvalues of the matrix A are
$$\mu_k^A = 1 + \lambda\theta 4 \sin^2(k\pi/2m) \geq 1\,, \quad k = 1, \ldots, m-1$$
So the symmetric matrix A is positive definite. Differentiating G w.r.t. x yields
$$G_x = Ax - \hat{b}\,, \quad G_{xx} = A\,,$$

which shows that A is the Hessian matrix of G. For $G \in C^2$ recall

G is strictly convex. $\iff$ The Hessian matrix of G is positive definite.

Hence the quadratic form G defined above is strictly convex and has a unique minimum on each convex subset of $\mathbb{R}^N$, $N = m-1$. Notice that $x \geq 0$ defines a convex set.

2.) Apply the Theorem of Karush, Kuhn, Tucker (KKT):

For the minimum x^* of a convex function G under the side conditions $H_i(x) \leq d_i$ for $i = 1, \ldots, N$, where H_i are convex, the following holds: There are $y_i \geq 0$ such that

$$\frac{\partial G(x^*)}{\partial x_j} + y_1 \frac{\partial H_1(x^*)}{\partial x_j} + \ldots + y_N \frac{\partial H_N(x^*)}{\partial x_j} = 0 \quad (j = 1, \ldots, N)$$

with $H_i(x^*) \leq d_i$ and $y_i(H_i(x^*) - d_i) = 0$ for $i = 1, \ldots, N$.
(references include [Stoer & Witzgall], [Strang])

In our application, $N = m-1$. Inequality $x \geq 0$ or $x_i \geq 0$ for all $i = 1, \ldots, m-1$ leads to $d_i = 0$ and $H_i(x) = -x_i$. The Theorem of KKT implies the existence of $y \geq 0$ with

$$\frac{\partial G}{\partial x_j} + y_j \frac{\partial H_j}{\partial x_j} + 0 = \frac{\partial G}{\partial x_j} - y_j = 0$$

and $y_i x_i = 0$ for all i. For our special G we conclude

$$G_x = Ax - \hat{b} \quad \Longrightarrow \quad Ax - \hat{b} - y = 0$$

Hence $(***)$ results from the KKT-Theorem and processes a unique solution. This then carries over to $(**)$.

B. Numerical Solution

A direct solution of $(**)$ is possible. Brennan & Schwartz suggest to proceed as follows:

> Solve $Aw = b$ *componentwise* such that
> the side condition $w \geq g$ is obeyed.

This is a somewhat vague outline of an approach: the implementation matters. It is based on the Gaussian elimination, which in its first phase transforms $Aw = b$ into an equivalent system $\tilde{A}w = \tilde{b}$, so that $\tilde{A}$ is a triangular matrix (here bidiagonal). Then, in the second phase, the above principle of Brennan & Schwartz can be solved with one loop. When $\tilde{A}$ is upper triangular, then this loop to *solve* $\tilde{A}w = \tilde{b}$ is a *backward* recursion. For a lower triangle $\tilde{A}$, the loop is *forward*. If in the ith step of the loop $\tilde{w}_i$ denotes the component of the solution of $\tilde{A}w = \tilde{b}$, then $w_i := \max\{\tilde{w}_i, g_i\}$ appears to be a realization.

But w depends on the loop's order. Only one direction works. An implementation must make sure that the characteristic structure of the option is matched. For a **put** this means:

Let $i_{\rm f}$ be the index of the node S_i that is closest to the contact point*, where V touches the payoff. Or more definitely, in the transformed variables,

$$w_i = g_i \text{ for } i \le i_{\rm f}\,, \quad \text{and}$$
$$w_i > g_i \text{ for } i_{\rm f} < i \le m.$$

This structure is known from the theory, but $i_{\rm f}$ is unknown. For the put, $w_1 = g_1$ is the starting point, and the $w_i := \max\{\tilde{w}_i,\, g_i\}$-loop is *forward*. Hence $\tilde{A}$ must be a lower triangle, which amounts to an RL-decomposition of A. This establishes the lower triangle $\tilde{A} := L$, and $\tilde{b}$ is solution of $R\tilde{b} = b$. The first components of the loop will be $w_i = g_i$, until the first index with $w_i > g_i$. This fixes the index $i_{\rm f}$.

Algorithm (put)

1st phase: Calculate the RL-decomposition of A. Then set $\tilde{A} = L$ and obtain $\tilde{b}$ from $R\tilde{b} = b$ (backward loop).

2nd phase: forward loop, start with $i = 1$. Calculate the next component of $\tilde{A}w = \tilde{b}$; denote it $\tilde{w}_i$.
Set $w_i := \max\{\tilde{w}_i,\, g_i\}$.

The **costs** are low (solution of a linear system with tridiagonal matrix). It can be shown that the above procedure for a standard option with the underlying matrix A works well.

For a **call** one proceeds the other way: The loop starts with $w_m = g_m$ and the second phase is a backward loop. To make this possible, in the first phase the "traditional" LR-decomposition of A establishes an upper triangle $\tilde{A} = R$, and $\tilde{b}$ is obtained from $L\tilde{b} = b$ in a foward loop.

Remark on the accuracy:

Since $V(S,t)$ fails to be twice continuously differentiable w.r.t. S at $S_{\rm f}(t)$, we expect some bad influence on the accuracy. (Recall that Crank–Nicolson even assumes $y \in C^4$.) But in spite of this lack of smoothness, the Crank–Nicolson approach here is sufficiently accurate, oscillations diminish rapidly. The lack of smoothness in the payoff is worse. Even extrapolation works rather well, although the assumptions of smoothness are not satisfied.

* provided the chosen interval is large enough, $S_1 < S_{\rm f}$

Outlook

This concludes the introduction in basic Computational Finance. An essential part of the course are the exercises (down-loadable), in particular the programming assignments.

If this is the material for one semester, then there will be more time left for some additional topics. (In my course, typically, there are two weeks left.) This additional material is not included in these course notes, because it will differ from course to course, depending on the interests and the knowledge of the students. For a textbook explaining additional material, see [Seydel: Tools for Computational Finance, Sixth Edition (2017)], from which we take the following section numbers. Possible topics include

- as further analytic method the integral equation of Kim (§4.8.4)
- *upwind* scheme and its relevance (§6.4 – 6.5)
- a penalty method in the two-dimensional case (§6.7)
- case studies, such as the two-dimensional tree method of Exercise 6.2
- jump diffusion (§1.9, §7.3)

If there is enough time for a two-semester course or an accompanying seminar, then one may address

- finite elements (Chapter 5)
- nonlinear Black–Scholes problems (§7.1 – 7.2)